INTERNAL CONTROL INFORMATION AND AUDIT PROGRAM REVISION: AN EMPIRICAL STUDY OF HUMAN JUDGMENT

By

RICHARD H. TABOR

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF
THE UNIVERSITY OF FLORIDA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORDIA

1980

ACKNOWLEDGEMENTS

I thank my supervisory committee, Dan Smith, Doug Snowball, Bill Messier, and Joe Reitz, for their guidance and support. Additional thanks go to Gary Holstrum, Charlie Swan, and others who contributed to the development of the materials used in this study.

I also am grateful to those auditors who gave of their valuable time to participate in this study and to the four firms for allowing their people to take part.

Finally, my thanks go to Teri Greene for contributing his typing skills to this project and to the members of my family and close friends who offered their support by continuously asking the important question: "Aren't you finished yet?"

TABLE OF CONTENTS

LIST OF TABLES

TABLE

LIST OF FIGURES

FIGURE

Abstract of Dissertation Presented to the Graduate Council
of the University of Florida in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

INTERNAL CONTROL INFORMATION AND AUDIT PROGRAM REVISION: AN EMPIRICAL STUDY OF HUMAN JUDGMENT

By

Richard H. Tabor

August 1980

Chairman: E. Dan Smith
Major Department: Accounting

This dissertation has provided additional evidence to facilitate a better understanding of the role of auditor judgment in the evaluation of IAC information and the decision as to the extent of subsequent audit procedures (substantive tests). A laboratory experiment was conducted to gather data which were used to examine auditor judgment about the reliability of an IAC system and the resulting selection of a sample size for a substantive test. The participants were 109 members of four "Big-Eight" accounting firms. A descriptive approach was taken within the framework provided by the Brunswik lens model to examine these decisions for a specific audit task.

Two key decision points within the audit process that require the exercise of professional judgment were examined in this study. These included: (1) the preliminary audit planning stage and (2) the

audit program revision stage. Examination of the responses from these stages resulted in descriptions of the auditors' judgments and the evaluation of auditor consensus and self-insight.

Responses of the auditors at both the preliminary audit stage and at the audit program revision stage suggest that, within the context of this study, the most important of three key controls in deciding that "recorded sales are for valid transactions" is the "control over physical shipment." The controls for "shipment authorization" and "credit approval" followed in importance in that order.

The extent of judgment consensus at the audit program revision stage, as examined through the application of canonical correlation and Pearson product-moment correlation, was found to be slightly lower than in other audit studies. The complexity of the audit task was thought to be a contributing factor to this finding. The nature of the audit task also was thought to contribute to the relatively low self-insight indices that were found. Firm differences and experience levels were found to be significant in examining the importance of controls, the extent of consensus, and the level of self-insight exhibited by the auditors.

In summary, although other studies have examined similar decisions made by auditors, some specific and significant limitations were thought to exist in these studies. The present study sought to overcome these limitations. Primarily, a more realistic setting was used, with a greater degree of control over the significant factors impacting upon the decisions of the auditors. This research has been presented with the hope that it can in some way serve as a stepping stone for future research in this vitally important area.

CHAPTER I

INTRODUCTION

The demand for the services provided by external auditors has increased in recent years as the complexities of the business world have increased. The knowledge and expertise of the independent auditor have resulted in a natural dependency on the auditor to provide information or services desired by management, interested third parties, or the general public.

As prescribed by the second standard of field work [AICPA, 1979a, Section 320], a primary area of interest and concern of the auditor when engaged to perform an audit is the functioning of the internal accounting controls (IACs). Internal control has been defined as follows:

> Internal control comprises the plan of organization and all of the coordinate methods and measures adopted within a business to safeguard its assets, check the accuracy and reliability of its accounting data, promote operational efficiency, and encourage adherence to prescribed managerial policies [AICPA, 1979a, Section 320.09].

This definition incorporates both accounting and administrative controls.[1] However, the extent of the auditor's concern for internal control depends primarily upon the type of engagement or audit objective. Recent reports relating to the evaluation of internal control (i.e.,

[1] See AICPA Professional Standards, Sections 320.09-320.13 for a discussion of administrative versus accounting controls.

The Foreign Corrupt Practices Act of 1977, Required Communication of Material Weaknesses in Internal Accounting Control [AICPA, 1979a, Section 323], and the Report of the Special Advisory Committee on Internal Accounting Control [AICPA, 1979b]) have publicized the significance of IAC evaluation to both management and external auditors.

The primary importance of IAC evaluation from the external auditor's perspective lies in the impact of the evaluation of the audit tests to be performed. As indicated in the AICPA Professional Standards:

> There is to be a proper study and evaluation of the existing internal control as a basis for reliance thereon and for the determination of the resultant extent of the tests to which auditing procedures are to be restricted [AICPA, 1979a, Section 320.01] (underlining added).

The present study examines this particular relationship within a specific audit setting.

This chapter introduces the concepts of IAC evaluation, compliance testing and substantive testing, followed by brief discussions of the relationships among these concepts and the importance of audit judgment to the audit process. The research objectives of the study are presented and discussed; a brief description of the research approach then is provided. Finally, the organization of the dissertation is presented.

Compliance and Substantive Tests

The extent of reliance on IAC is determined by evaluation of the results of compliance testing. Compliance testing is used to provide reasonable assurance that the control procedures within a

company are operating as prescribed through preliminary evaluation. The results of compliance testing provide the necessary information to determine the nature, timing, and extent of substantive procedures.

Substantive testing includes those procedures necessary to obtain sufficient competent evidential matter as prescribed in the AICPA's third standard of field work [AICPA, 1979a, Section 330]. These procedures include: (a) tests of details of transactions and balances and (b) analytical review procedures applied to financial information. Therefore, it is a combination of compliance test procedures and substantive test procedures that constitutes the main component of the audit program that is to be followed.

Auditor Judgment

As a result of the additional emphasis on generating IAC information, we might expect to find a redistribution of resources by the auditor with regard to the appropriate combination of audit procedures, or higher audit fees for the additional work that would be required. From a research perspective, there is an obvious need for additional examination of the auditor's decisions that result from the exercise of his professional judgment. Specifically, his incorporation of the important inputs of preliminary IAC evaluation and compliance test results into a decision regarding the optimal extent of additional audit procedures (substantive tests) warrants examination.

Professional judgment is fundamental to the auditing profession and pervades the entire audit process. Many statements have been made regarding the necessity to exercise "judgment" in evaluating both the qualitative and quantitative considerations in an audit

situation. Some statements found in the AICPA Professional Standards that relate specifically to this study include the following:

> In the observance of generally accepted auditing standards, the independent auditor must exercise his judgment in determining which auditing procedures are necessary in the circumstances to afford a reasonable basis for his opinion. His judgment is required to be the informed judgment of a qualified professional person [AICPA, 1979a, Section 110.04] (underlining added).
>
> ...the purpose of tests of compliance with accounting control procedures is to provide "a reasonable degree of assurance that they are in use and are operating as planned." What constitutes a "reasonable" degree of assurance is a matter of auditing judgment,... [AICPA, 1979a, Section 320.60] (underlining added).
>
> Although statistical sampling furnishes the auditor with a measure of precision and reliability, statistical techniques do not define for the auditor the values of each required to provide audit satisfaction. Specification of the precision and reliability necessary in a given test is an auditing function and must be based upon judgment in the same way as is the decision as to audit satisfaction required when statistical sampling is not used [AICPA, 1979a, Section 320A.03] (underlining added).
>
> The amount and kinds of evidential matter required to support an informed opinion are matters for the auditor to determine in the exercise of his professional judgment after a careful study of the circumstances in the particular case [AICPA, 1979a, Section 330.09] (underlining added).

The fact that professional judgment plays a major role in the decision process of auditors has been recognized for many years. Mautz [1959, p. 44] presented conclusions concerning auditor judgment as follows: (1) judgment must play a major role in auditing, and (2) auditors would do well to recognize this and acquaint themselves with the process of judgment formation.

Although the following statement was made in reference to accounting in general, it is particularly germane to the auditing process and the specific purpose of this study:

> Judgment is, of course, a vital part of any professional's work. In accounting it plays an important role every step of the way. But that does not mean that it is a mysterious process, undefinable and inexplicable. We know that the processes that feed professional judgment are varied and complex, yet it does not follow that we cannot make some progress in their analysis and description [Bernstein, 1967, p. 9].

Recently, there has been increased recognition that auditor judgment must be further evaluated and its impact understood. A number of recent research publications, e.g., Kennedy [1977], Gibbins [1977], and the Journal of Accounting Research Supplement on Human Information Processing [1976], have presented discussions of recent studies that concern auditor judgment. This study also emphasizes the importance of auditor judgment as recognized in the specific research objectives discussed in the following section.

Research Objectives

The objective of this study is to examine and describe (within a specific audit setting) the impact of internal control information (including the specific results of compliance tests) on the auditor's judgment concerning the reliability of the IAC system and on his resulting decision as to the extent of substantive testing. This objective is pursued through the use of a laboratory experiment which utilizes auditors from four of the "Big-Eight" public accounting firms. Questions relevant to this particular objective and addressed in this study can be summarized as follows:

(1) To what extent do auditors reflect consensus in terms of their specification of the reliability of a client's internal controls (both prior to and following the evaluation of compliance tests) and to what extent is there consensus at the later point in the audit process where the question as to the appropriate amount of substantive testing to perform is addressed?

(2) How do auditors respond to the results of compliance testing in terms of deciding the appropriate amount of substantive testing, when they are presented with results that may confirm or disconfirm their original beliefs about the existence and effectiveness of key controls?

(3) How important are the various key internal controls in arriving at the appropriate amount of substantive testing as determined by: 1) the confidence levels chosen for compliance testing, 2) the subjective weights assigned by auditors, and 3) weights determined from the analysis-of-variance (ANOVA) technique? And, to what extent do auditors exhibit self-insight in understanding the relative importance of the key internal controls?

(4) Do such factors as firm differences and experience levels have a significant effect on the decisions made by auditors?

Judgment Consensus

Consensus is reflected by agreement among auditors when using the same data to reach a decision. Lack of consensus, or disagreement among auditors, is thought by many to be costly. For example, Joyce has stated this concern as follows:

> Within firms, the existence of the review process suggests that individual differences exist and are likely to be costly if unresolved. The increasing concern within the profession about "quality control" issues . . . The existence of continuing education programs within at least the major audit firms is further evidence of their willingness to consume resources to restrict judgment variance among their professional staff [1976, p. 31].

This study is expected to provide additional evidence as to the extent of agreement among auditors when provided with a particular audit situation. Although the purpose of this study is not to provide any information as to the costs of any differences that may be found, additional information as to the source of such differences would be a meaningful contribution. The study evaluates decisions made by the auditors at both a preliminary decision point in the audit process (prior to conducting compliance tests) and, more extensively, at the audit program revision stage (after evaluation of compliance test results).[2] This analysis may provide insights into the point at which disagreement among auditors begins to emerge.

---

[2]See Chapter II above for a discussion of these various stages of the audit process.

Evaluation of Evidence

The independent auditor's objective is to obtain sufficient competent evidential matter to provide him with a reasonable basis for forming an opinion under the circumstances. He is also expected to be thorough in his search for evidential matter and objective in its evaluation [AICPA, 1979a, Section 330.09-330.15]. However, very little is known concerning the auditor's reaction to, and incorporation of, the evidence he has gathered.

Of primary interest in this study is the auditor's reaction to evidence provided by IAC information. The results of this study should provide insights into the auditor's evaluation of information that has been lacking in previous studies. This opportunity is enhanced by the requirement that the auditor use a statistical sampling approach in evaluating the evidence concerning the key IACs within a controlled audit environment.

Importance of Controls and Self-Insight

The fact that the auditors will be presented with a specific audit objective and the key controls to be considered in meeting the objective provide an excellent opportunity to evaluate the relative importance of the controls. There are three possible methods for determining the relative importance of the controls. First, confidence levels used in determining sample sizes for the compliance tests can be used to infer the relative importance of the three IACs. Second, the ANOVA technique provides a measure of importance from the evaluation of the decisions made by the auditors after their evaluation

of the compliance test results. Finally, a post-experiment question requires a subjective weighting of the relative importance of the three key controls.

The indications of the relative importance of the controls allow for additional evaluation of a measure of consensus and also for a calculation of a self-insight index. This ability of the auditor to estimate the relative importance he places on the cues (controls) in making his judgments is considered most important within the auditing profession. Joyce, in discussing the importance of self-insight to auditors, made the following statement:

> One of the implications of poor self-insight in situations where a considerable amount of professional expertise is communicated verbally "on the job" (between senior and junior accountants on an audit, for example) is clear: A distorted representation of one professional's decision behavior will be transmitted to another professional [1976, p. 52].

This study will provide evidence that may aid the assessment of the extent to which the lack of self-insight represents a significant problem within the auditing profession.

Firm and Experience Effects

This study will also provide information as to the effect of two key factors on the judgments made by the auditors. Of specific interest are differences that may be attributable to differences in the firm affiliation of auditors and/or differences between various experience levels.

Research Approach

Past research concerning the audit process has been both of a normative and a descriptive nature. This study will take a descriptive approach. Support for this type of research is provided by Kaplan:

> It seems difficult to make a central attack on the audit process because so little is known about what actually constitutes a good audit. This suggests that, rather than start with research on normative models in an attempt to improve existing audits, we devote significant resources to developing descriptive models of what auditors are actually doing now [1977, p. 9].

Further support for this approach is provided by recent calls for additional research efforts in the area of audit judgment by Kaplan [1977] and Libby and Lewis [1977].

Recently, accounting researchers have adapted the "lens model" approach from psychology to the examination of judgments made in accounting. This particular approach appears most consistent with the descriptive nature of this study. The appropriateness of the lens model for a study such as this has been recognized by Libby and Lewis:

> . . . this approach (lens model) is particularly useful in studying the impact of information set variables on decision rule form, stability or learning, cue usage, and decision accuracy, reliability, and predictability or in descriptive studies of these variables. This method of modeling judgment provides a compromise between the overly simplified approach of asking subjects to describe the weights they place on information and the more complex and expensive process tracing models that have been used in the study of judgment [1977, p. 248] (underlining added).

A more extensive discussion of the "lens model" methodology is presented in Chapter III.

The "lens model" will be used in this study to examine the auditors' decisions with respect to the evaluation of internal control information. More specifically, the audit program revision stage of the experiment is arranged in a $2 \times 3 \times 2$ factorial design which results in twelve case situations presented to each auditor. The analysis of the responses is discussed in Chapter III. The contents of the remaining chapters are summarized below.

Dissertation Organization

This dissertation is composed of five chapters. Chapter II presents a more detailed discussion of the audit process and the roles of statistical sampling and professional judgment at various stages of the process. A better understanding and description of the audit process enables a focus on the key decision points which are of interest in this study. Also, the roles of statistical sampling and professional judgment are evaluated and the complementary nature of their relationship is established. Relevant literature reviews for these areas are presented and discussed throughout Chapter II.

The general research design and methodologies used are described in Chapter III along with the details of the experimental setting and the materials utilized. Characteristics of this study representing both strengths and limitations are also noted in this chapter.

Chapter IV presents the results of the study according to the objectives presented previously. The final chapter presents implications of the results and some ideas for future research.

CHAPTER II

THE AUDIT PROCESS AND PROFESSIONAL JUDGMENT

Introduction

The problem of interest has been broadly stated as a lack of understanding of the impact of IAC information on other audit procedures. The sequential nature of the auditing process and the fact that professional judgment plays a major role in the evaluation of information are both contributing factors to this problem. The purpose of this study is to further describe the impact of IAC information on the auditor's judgment concerning the reliability of the IAC system and his resulting choice of the amount of substantive testing.

In this chapter the relevant components of the audit process and the role of professional judgment are introduced and discussed. Relevant literature in these areas will also be noted and in some cases discussed.

The Audit Process

Figure 2-1 presents a diagrammatic representation of the audit process. Many studies in auditing have dealt either with one of these steps exclusively or various combinations of steps. This study is concerned with the decisions required at the two key points represented by Step 3 and Step 5, with primary interest in the responses of the auditors at Step 5.

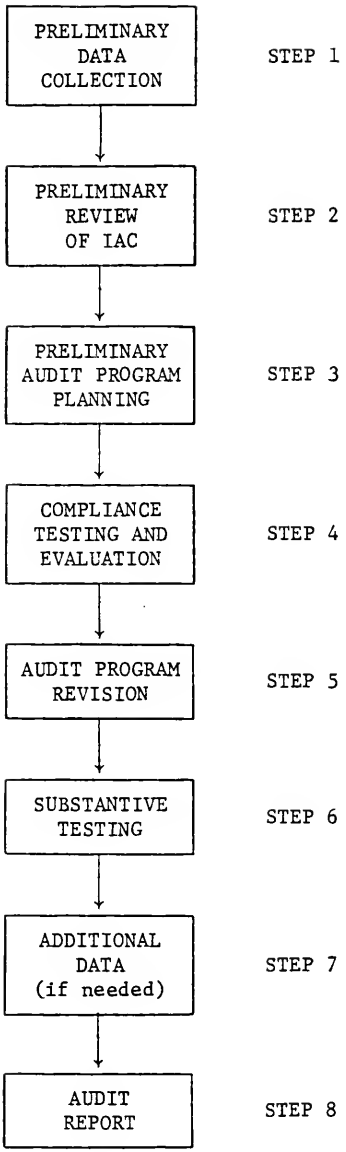


Figure 2-1. The audit process.

Both of these two key decision points within the audit process call for professional judgment on the part of the auditor. First, as a result of internal control evaluation,[1] the auditor must decide on the degree of reliability he associates with existing controls in order to develop the appropriate preliminary audit program (Step 3). This program would include both compliance and substantive tests. Second, and more importantly,[2] the auditor may be required to adjust his planned substantive tests (Step 5) as warranted by his evaluation of the results of compliance testing (Step 4).

One study [Ashton, 1974] has particular relevance to the present study since it deals specifically with the auditor's ability to assess the internal control environment (Step 2) and the extent to which auditors agree at this point in the audit process. Other studies [Joyce, 1976; Mock and Turner, 1979] have expanded the decisions required by their subjects to include some type of program planning (Step 3), such as a specification of either man hours or sample sizes.[3] The present study will extend to the audit program revision stage (Step 5) of the audit process.

The dynamic nature of the audit environment requires the auditor to adjust his plans in response to additional information. The

---

[1]This would include the completion of the internal control questionnaire, walk-through of the system and its evaluation.

[2]The second decision point is more important because actual allocation of the firm's resources results from this decision.

[3]These three studies will be discussed in greater detail later in this chapter.

auditor will plan the nature, timing, and extent of the audit procedures he deems most appropriate as a result of his knowledge about the client[4] and the specific objectives of the audit. The auditor is then required to evaluate additional evidence to determine if a revised audit program is appropriate. More specifically, additional information obtained through the performance of compliance tests on internal controls and the resulting evaluation of such tests must be considered.

The importance of analyzing the program revision stage of the audit process was recognized by Joyce as follows:

> The selection of audit program planning as a measure of audit work to perform can be criticized on the grounds that auditors may disagree on an audit program yet end up performing the same tests and arriving at the same opinion. Differences in the initial audit programs might vanish as they are revised in view of information collected as the audit progresses. I am unaware of any empirical evidence to support or refute this contention [1976, p. 35].

Although individuals and accounting firms may express different audit philosophies and may prefer a different emphasis on audit procedures, the essence of the audit process as reflected in Figure 2-1 is generally accepted. That is, evidence is gathered in a sequential process that allows for the nature, extent, and timing of such evidence to be controlled in a manner that is deemed optimal. It is the evidence gathering and evaluation process that ultimately results in the decision as to the optimal allocation of resources within an audit. Therefore, this decision should be a major concern of those performing the audit. The importance of the recognition of the

---

[4]This would include any information he has obtained personnally in either the current or previous years and the information made available to him from the working papers relating to his client.

sequential nature of the audit process has been highlighted by Wright [1976]. A focus of this study will be on the sequential nature of the audit process as the auditor uses his professional judgment in making the decisions relative to Step 3 (preliminary audit program planning) through Step 5 (audit program revision).

The importance of understanding the revision stage of the audit process has been expressed by Mock and Turner as follows:

> Understanding the auditor's decision making process may lead to decision aids which will assist the auditor in evaluating audit evidence. It certainly seems unlikely that significant improvements will be forthcoming without some general agreement on how auditors act when faced with decisions on how much substantive evidence is appropriate in different internal control situations [1979, p. 277].

Internal Control

The importance of internal control evaluation and its impact on the ultimate issuance of an audit opinion has led to many recent research efforts relating specifically to internal control. For example, Burns [1974], Burns and Loebbecke [1975], and many others have discussed internal control evaluation from various perspectives. Mathematical approaches to the analyzing and/or modeling systems of internal control have been presented by Grimlund [1978], Bodnar [1975], Cushing [1974], Yu and Neter [1973], and others.

A certain level of internal control is expected and even required to be present within a firm (see The Foreign Corrupt Practices Act of 1977). The accumulation of additional audit evidence is affected directly by the IAC environment. Of interest in this study is the relationship between internal control evaluation and subsequent

compliance and substantive testing. The impact of IAC evaluation on audit scope adjustments has been discussed by Morris and Anderson [1976] and Smith [1972]. This type of analysis was greatly extended by Kinney [1975a; 1975b]. Kinney presented a decision theory approach to examinations of the relationship of IAC evaluation and compliance and substantive tests that consider such relevant factors as costs of sampling, costs of errors, and utility functions of the auditor. However, the nature of the decision theory approach extends beyond the scope of the present study.

Within the context of the present study, the relationships between the key factors of IAC evaluation and substantive testing can be related to the concept of risk and summarized as follows:

$$(1-R_o) = (1-R_c) \cdot (1-R_s)$$

where, $(1-R_o)$: Total overall risk [R_o = overall reliability]

$(1-R_s)$: Risk of substantive tests [R_s = reliability of substantive tests]

$(1-R_c)$: Risk of internal control [R_c = reliability of internal control]

The use of statistical sampling has been advocated as one way of controlling for overall audit risk [Roberts, 1978; Warren, 1979].

Statistical Sampling and Professional Judgment

The nature of the evaluation process performed by auditors and the role that the concept of risk has within the decision process has led to an increasing appreciation by auditors of statistical sampling techniques.

Elliot and Rogers [1972] discussed in a very basic fashion the use of statistical sampling in auditing. Arkin [1976] analyzed the use of statistics in auditing with particular emphasis upon the area of internal control compliance. Other discussions of statistical sampling in auditing include Loebbecke and Neter [1975] and Teitlebaum and Robinson [1975].

Regardless of the extent to which statistical sampling techniques are employed, the importance of professional judgment is not diminished. The use of statistical sampling for planning compliance and substantive tests would require judgments such as the following: 1) expected error rates in the population, 2) desired level of confidence, 3) desired precision level, 4) standard deviation of population items, and 5) the amount considered material. The role of judgment in statistical sampling has been evaluated by Boer [1974]. Uecker and Kinney [1977] reported results of a study to investigate the type and severity of errors that practicing CPAs make in the judgmental evaluation of statistical sampling outcomes. Results indicated that the problem of subjectively evaluating sample outcomes may be significant. In their study, seventy-four percent of the CPAs made at least one serious error of judgment and fifty-six percent made at least two such errors.

The role of professional judgment becomes paramount as statistical sampling techniques must be integrated with various stages of the audit process. The most obvious integration occurs within internal control assessment. That is, the auditor must decide the extent to which he will rely upon internal controls in planning the nature, timing, and extent of compliance and substantive tests. Due to the

difficulty in dealing with all of these planning variables, this study will emphasize the decisions concerned with the extent of testing.

Given the role of professional judgment within the audit process in determining the extent of audit work to perform, the commitment of a firm's resources could vary considerably with variations in auditor judgments. In the absence of an operational normative model to rely upon, a strict evaluation of such judgments as being "correct" or "optimal" is not possible. The appropriate evaluation criteria then becomes the degree of consensus among auditors when presented with the same audit environment and audit evidence. Einhorn [1974] suggests that consensus among experts is one of several necessary, although not sufficient, conditions for the existence of professional expertise.

Although the actual cost associated with a lack of consensus among auditors can not be measured directly, various forms of evidence are available that suggest the auditing profession strives for consensus. Such factors include the following: 1) the administration of a uniform CPA exam, 2) issuance of standards to follow in the practice of auditing, and 3) the use of training schools by individual firms to guide the learning process of their employees. Given the actions taken to strive for consensus by the auditing profession, an inference can be made that the cost of such actions must be less than the alternative cost of not having a certain degree of consensus.

A major question of interest then becomes, "where within the audit process should this consensus be reached?" Should consensus exist at every step of the audit process, or is agreement as to the

opinion that is to be issued the only point at which consensus is important? The answer to this question seems to be that while the issuance of the opinion is a key point of consensus, firms are also interested in the optimal utilization of their available resources as they seek to reduce the probability of making an incorrect decision.[5] The lack of consensus at decision points associated with the commitment of a firm's resources could result in a less than optimal use of such resources. The topic of consensus and other topics in the human information processing area that are relevant to this study are discussed in the following section.

Human Information Processing

Psychological Literature

The area of human information processing (HIP) has been examined extensively in the psychological literature. Three broad approaches to the study of HIP include: 1) Bayesian, 2) regression, and 3) cognitive complexity. A number of authors have discussed the relative merits of the different approaches to study HIP [Slovic and Lichtenstein, 1971; Slovic, Fischhoff, and Lichtenstein, 1977] with comments concerning the appropriate use of each of these approaches.

A Bayesian approach is most concerned with the evaluation of an "optimal" manner to combine information, or "how men _should_ think."

---

[5]The auditor is concerned with controlling the probability of accepting the financial statements as being "fairly presented" when, in fact, they are not. The cost of this Type II(3) error could be very large.

Thus, it provides a normative model for the evaluation of probabilities expressed by judges or subjects. A cognitive complexity approach to the evaluation of HIP requires that individuals be put into complex situations. The processing of information is then evaluated as decisions are made to determine the "style" used.

This study is not concerned with the ability of the subject to "optimally" process information (Bayesian approach) or the style used (cognitive complexity approach). Therefore, the regression approach is considered most appropriate for use in this study, with the "lens model" considered the most applicable "representative design" within the regression approach.

The lens model[6] is based on the assumption that man makes inferences about the state of an uncertain environment on the basis of uncertain information. Therefore, the outcomes that result from an interaction between man and his uncertain environment are most certainly probabilistic. The lens model specifies a relationship between the probabilistic outcomes of the environment and the prediction or judgment by an individual of a particular state of that environment.

Analysis of the information represented on the "right-hand-side" of the lens model (the actual and predicted subjects' responses) can result in the capturing of the decision policy of the subjects. The correlation of the actual responses between the subjects is considered an appropriate measure of consensus (agreement). Also, the mathematical modeling of the judgment process allows for the evaluation

---

[6]For a more complete presentation and explanation of the lens model refer to Chapter III, Brunswik [1949], or Slovic and Lichtenstein [1971].

of the relative importance of the various cues (factors). Self-insight indices can then be obtained by correlating the mathematical weightings of the cues with the subjective weightings provided by the subjects.

The following sections present a discussion of the evaluation criteria of consensus and self-insight that were mentioned above. These particular areas have been investigated in the psychological literature and have the most relevance to this study.

Consensus. Consensus is the extent of agreement among judges using the same information at the same point in time. Many psychological studies have indicated that consensus among judges is very poor (e.g., Goldberg, 1968): most accounting studies have indicated a relatively high degree of consensus.

The extent of consensus found in most studies is thought to be a function of the judgment task and the specific environment. However, most realistic tasks have yielded results which indicate a lack of consensus among persons considered professionals with respect to the judgmental task. Various opinions exist as to the source of such differences. For example, Ashton and Kramer state the following:

> Some psychologists (e.g., Brehmer, 1976; Dawes, 1975; Edwards, 1971) have argued recently that certain formal or statistical characteristics of the judgment task, and not the substantive or content characteristics of the task nor the particular subjects in the task, are the most important determinants of human information processing and decision making [1980].

Self-insight. Self-insight is the ability of the judge to determine the relative weights of the various factors he uses in his judgment process. Studies have indicated that judges exhibit poor

insight into their own weightings of factors [Slovic, 1969; Slovic, Fleissman, and Bauman, 1972]. Researchers have found differences between the subjective weights provided by judges and those computed by a mathematical model of their judgmental policies. For example, the results from psychological studies have indicated that judges overestimate the less important factors and underestimate the more important factors (e.g., Slovic, 1969).

Of additional interest is the finding that non-linear models of the judgment process contribute little as compared to linear models [Goldberg, 1968; Slovic, 1969; Slovic and Lichtenstein, 1971; Dawes and Corrigan, 1974]. The implication is that a simple linear model will satisfactorily approximate the manner in which an individual processes information. The adequacy of the linear model also will be examined in the current study.

Auditing Literature

Auditing researchers also have developed an interest in HIP, the applicability of the lens model approach, and the resulting model of the judgment process. The probabilistic relationship between man and his uncertain environment, as reflected in the lens model, is certainly present in the audit environment.

The auditor is required to evaluate the uncertain audit environment to ascertain that either "acceptable" or "unacceptable" situations exist. He must then make decisions concerning specific combinations and the extent of audit procedures that are most appropriate given his current beliefs about the state of the environment.

Within the specific judgmental context of this study, the true state of the environment is that either a specific control objective is "acceptably" achieved, or it is not. Of interest in this study is the extent of evidence required by the auditors to determine whether an objective has been "acceptably" achieved. Because the "true" state of the environment cannot be determined,[7] the analysis of this study is limited to the "right-hand-side" of the lens model where the auditor's judgment and his use of cues can be evaluated. A better understanding of this approach is obtained through a discussion of some recent studies in auditing.

Most Relevant Prior Research

Various studies have evaluated the extent of consensus among auditors at different points in the audit process. Results indicate that the degree of consensus among auditors apparently varies according to the particular stage in the audit process or according to the type of question addressed. Ashton [1974] and others who have replicated and extended his original study, Joyce [1976], and Mock and Turner [1979], provide a basis for summarizing the judgment literature in accounting. These studies will be discussed separately, with particular emphasis on their relationship to the present study.

Ashton Study, Replications and Extensions

Ashton was most concerned with the extent of inconsistency shown by auditors in their evaluation of internal control. By

---

[7] For relevant discussions of the importance of the "true" or criterion value in the lens model analysis, see Casey [1976] and Ashton [1976].

presenting sixty-three practicing auditors with thirty-two cases (each a different combination of Yes/No responses to six questions found on an internal control questionnaire for payroll), judgments of internal control strength were elicited. These judgments were made on a scale from one (extremely weak) to six (adequate to strong).

Ashton used a descriptive analysis-of-variance (ANOVA) technique to evaluate two types of consistency. These are: 1) consensus, which refers to consistency across auditors at the same point in time, and 2) stability, which refers to consistency over time for the same data. By correlating the judgments of each auditor on the cases, both with the other auditors and with later administrations for the same auditor, Ashton found that the judgments exhibited a "fairly high level" of consensus ($r = .70$), stability ($r = .81$), and self-insight ($r = .89$).

By addressing the very crucial question of "to what extent do auditors agree in their assessment of the strength of an internal control questionnaire?" Ashton's study represents a major contribution to auditing research. Ashton chose to replicate his study with the use of upper division auditing students at The University of Texas at Austin [Ashton and Kramer, 1980]. These results were inconclusive with respect to whether the auditing students were good surrogates for practicing auditors. However, the results did indicate a similar high degree of consensus on the part of the students ($r = .66$), with auditors also showing a higher degree of self-insight ($r = .89$) as compared to the students ($r = .77$).

Ashton and Brown [1980] extended Ashton's earlier work through the addition of two cues to the six that had been used in the original study. The subjects used in this study were 31 practicing auditors from four offices of seven of the "Big-Eight" public accounting firms. A similar consensus index was found ($r = .67$), with self-insight indices averaging .86. It should be noted that this lower average self-insight index was obtained from auditors who were generally less experienced than those who participated in the original study, but who were of course more experienced than the students tested in the Ashton and Kramer study.

Hamilton and Wright [1977] also used the payroll environment to evaluate audit judgment in a modified replication of Ashton's original study. Seventeen practicing auditors from the same city were presented with five internal control questions relative to payroll. The average values of consensus and self-insight were .66 and .87, respectively. Interesting results with respect to experience levels showed that the group with more experience had both greater insight (.93 to .84) and greater consensus (.78 to .62). These particular findings are consistent with the results of the Ashton and Ashton and Kramer studies.

The most recent extension of Ashton's study of internal control relative to payroll was provided by Reckers and Taylor [1979]. Thirty practicing auditors were presented with a completed payroll questionnaire containing responses to approximately 36 questions; these responses were varied to form 5 different cases. In analyzing the subjects' responses as to the reliability of the internal control system (0-100 percent), an average inter-rater correlation of .155

was obtained. This increased slightly (to .357) for those participants with more than seven-and-one-half years' experience. These results suggest that the level of consensus is reduced as greater realism is introduced into the audit task. When auditing professors were presented with the same task similar results were obtained ($r = .28$). However, Ashton [1979b] expresses some concern about the cases used and assumptions made by Reckers and Taylor.

The studies discussed above provided a meaningful evaluation of audit judgment within the payroll environment and with respect to the strength of an internal control environment. However, because of their nature, extensions of these studies appear warranted. A natural progression of auditor judgment research would involve a shift in focus to a different point in the audit process. This shift appears necessary since the conclusion that auditors have exhibited a "fairly high level" of consensus relative to the strength of internal control says nothing as to the extent of their agreement at a later point in the audit process; or more specifically, as to how this level of agreement extends to decisions about the extent of other audit procedures to perform. Joyce provided such an extension.

Joyce Study

Joyce [1976] extended the research in the area of auditor judgment by focusing upon the decision regarding the extent of audit work (planned man-hours) rather than upon the assessments of the quality of internal control. This represented an analysis of auditor judgment for a different question at a later point in the audit process.

Thirty-five practicing auditors were presented with a set of sixteen systematically varied combinations of stimulus information related to an audit program for accounts receivable.[8] The auditors were asked to indicate the planned extent (in man-hours) of five audit procedures.

Some conclusions resulting from Joyce's use of correlation measures and ANOVA (or MANOVA) can be summarized as follows:

(1) There was considerable disagreement concerning how the independent variables (factors used in determining the extent of audit procedures) should be weighted.

(2) There was little consensus ($r = .37$) among auditors concerning how much time should be planned for the audit procedures.[9]

(3) Main effects accounted for virtually all the reliable judgment variance.

(4) The level of consensus decreased as experience increased.

The stimulus information provided by Joyce was presented as having been developed from various sources such as an internal control questionnaire and analytical review data (i.e., significant ratios and turnovers). An appropriate adjustment to this experimental environment would be to provide the auditors with a more realistic setting and information as to the results of compliance tests on the internal

---

[8] Joyce actually performed two experiments, with the second experiment being a full replication of the 2^5 factorial conducted as Experiment 1. Joyce was therefore able to assess the impact of two-way and three-way interactions.

[9] Note that although Joyce's subjects made multiple judgments on each case, the consensus measure was calculated by summing the audit hours on each case and then correlating the subjects' judgment.

control information. Mock and Turner performed a study with these characteristics.

Mock and Turner Study

Mock and Turner [1979] sought empirical evidence about the effect of changes in internal controls on auditors' decisions on the extent of substantive testing. Seventy-one audit seniors and two audit supervisors (all from a single firm) were given information on improvements in internal controls from the previous year and asked to adjust the sample sizes for four specific audit procedures from the planned audit program. The improved internal controls were evidenced primarily by a general change in compliance test results of specific controls for the current year as compared to the previous year.

Some of the major results of the study include the following:

(1) Subjects reduced their judgmental samples in every case except one. (The subjects were dealing only with improvements in controls and the exception was probably due to the fact that the planned sample for this case was knowingly set quite low.)

(2) Subjects reacting to strong controls consistently recommended smaller sample sizes for planned substantive tests than did subjects reacting to fair controls.

(3) There was considerable variability in auditors' decisions. (Results showed both wide ranges and large coefficients of variation for all control procedures with sample size decisions for strong controls varying much less than for fair controls.)

(4) No evidence was found of a relationship between the subjects' backgrounds and their judgments.

(5) Possible effects of anchoring[10] were found. Those who had anchors recommended smaller sample sizes and mentioned in rationale memos that they gave attention to the previous sample sizes.

The present study is intended to be an important extension of the study by Mock and Turner. The decision to build upon and to extend their study is, to some extent, based upon their use of a realistic setting. Accordingly, some specific characteristics of Mock and Turner's experimental environment and important extensions and differences relative to this study are provided below. These particular characteristics of Mock and Turner's study may be summarized as follows:

(1) They provided general evidence about changing internal control strengths (resulting from compliance tests) on a year-to-year basis.

The present study provides more specific information pertaining to the results of compliance testing of key internal controls as compared to the preliminary evaluation of the current period. These specific results will then serve as a basis for adjusting the planned extent of audit procedures for the current period. This restriction will serve to reduce the possibility of incorporating confounding variables into the experimental setting, through superior control over the factors being considered by the auditor.

(2) They required the subjects to deal with a "perceived" audit objective in a broad audit setting (i.e., the Revenue Cycle).

---

[10]"Anchoring" is the process of adjusting from initial values or starting points to yield final estimates. See Tversky and Kahneman [1974] for a discussion of this heuristic principle.

This study also uses the Revenue (Sales) Cycle as the experimental environment. However, a specific audit objective is provided to the subjects and the internal control information is then keyed to this particular objective. This will allow the auditors to evaluate the evidence provided to them on a "common" basis and therefore make the comparison of their decisions more valid.

(3) The subjects were told the degree of reliance to place on internal controls for either fair or strong controls.

The subjects in this study are provided with the specific results of compliance tests (noncompliance rates for the sample sizes selected by the subjects) and must decide their degree of reliance on the key internal controls after receiving these results. This adjustment incorporates additional realism into the setting and provides greater assurance as to the message that is being provided to the auditor. The auditor can then interpret this message as he desires.

(4) The subjects were presented only with <u>improvements</u> in internal controls and asked to adjust a <u>planned</u> sample size.

This study provides various combinations of "confirming" and "disconfirming" evidence as to the existence and effectiveness of the key internal controls as compared to certain expectations. The subjects also are asked to adjust their own planned sample size for a substantive test rather than to adjust a planned sample size that has been provided to them. As this study is most concerned with the final sample size decision, it was felt that providing a planned sample size would influence the auditors' decisions. Again, a confounding factor would be introduced because it would not be clear what effect this

would have on each auditor. It was considered that an inferior measure of consensus among auditors would be obtained if a planned sample size was provided.

(5) The subjects were presented with only one combination of controls to use in planning a substantive program.

The subjects in this study are provided with twelve combinations of compliance test results for three key internal controls that are considered most important in meeting the specific objective. This is necessary to allow for the desired analysis relative to consensus and the weightings of the key variables. The single case approach taken by Mock and Turner did not allow for a within-subject evaluation.

(6) The subjects were required to make "judgmental" sample size decisions.

All sample sizes in this study are chosen from statistical sampling tables provided to the subjects. For compliance tests, tables were provided that gave the subjects various confidence levels and desired precision levels from which to choose. For the substantive test, a table that was derived from a difference estimation approach was provided; in using this table it is necessary for the subject to decide upon a confidence level and an amount for materiality. This approach provides additional information as to which factors were considered most important and/or altered as the auditor moved from case to case.

The above discussion provides a comparison of this study with the most recent study of Mock and Turner. Perhaps the most obvious intention of this study is to provide the auditor with a more compact audit environment and specific audit objective to a greater degree

than in previous studies. At the same time, the desire to present a realistic setting is also of concern.

Summary

The review of the psychology and auditing literature presented in this chapter provides additional support for the contention that the application of professional judgment, both in general and within the auditing context, deserves further investigation. Support is also provided for the importance of the evaluation of internal accounting controls and their role within the audit process.

Various deficiencies in previous auditor judgment studies were presented and discussed. General characteristics of this study were presented as significant improvements over prior studies. Specific details of the experiment that was conducted are presented in Chapter III in conjunction with a presentation of the methodology and experimental design.

CHAPTER III

THE EXPERIMENT

Introduction

The purpose of this chapter is to present the methodology and experimental design used in this study, along with the details of the experiment. First, in order to provide a better understanding of the descriptive approach used in this study, a brief description of the Brunswik lens model and the analysis of variance model is presented. A discussion of the specific task, the experimental design, and the methodology follow in the next section. Finally, the steps in the administration of this study and some expectations as to results are presented, followed by several recognized limitations of the study.

The Lens Model

The lens model as developed and described by Egon Brunswik [1943, 1949] has become an acceptable and popular model to use in human information processing (HIP) research. Because of the descriptive nature of this study, the lens model provides the most appropriate framework for the description and examination of the auditor's judgment process.

The lens model is considered a significant development within a "correlational" paradigm of the regression approach to the study of

human information processing [Slovic and Lichtenstein, 1971]. Brunswik's model emphasizes "the probabilistic interrelations between organismic and environmental components of the judgment situation" [Slovic and Lichtenstein, 1971, p. 655] as opposed to a focus on the judge (information processor).

Specification of the Univariate Lens Model

As shown in Figure 3-1, the univariate lens model utilizes both environmental and subject response data. The left side of the model represents the environmental system, while a subject's (decision maker's) actual and predicted responses are represented on the right side. Correlational measures are indicated for the relationships within and between those systems. The variables shown in the lens model are defined as follows:

X_i = the variables found in the information set which serve as cues to the subject.

Y_e = the criterion value or "distal variable." This value represents the actual result of an environmental event.

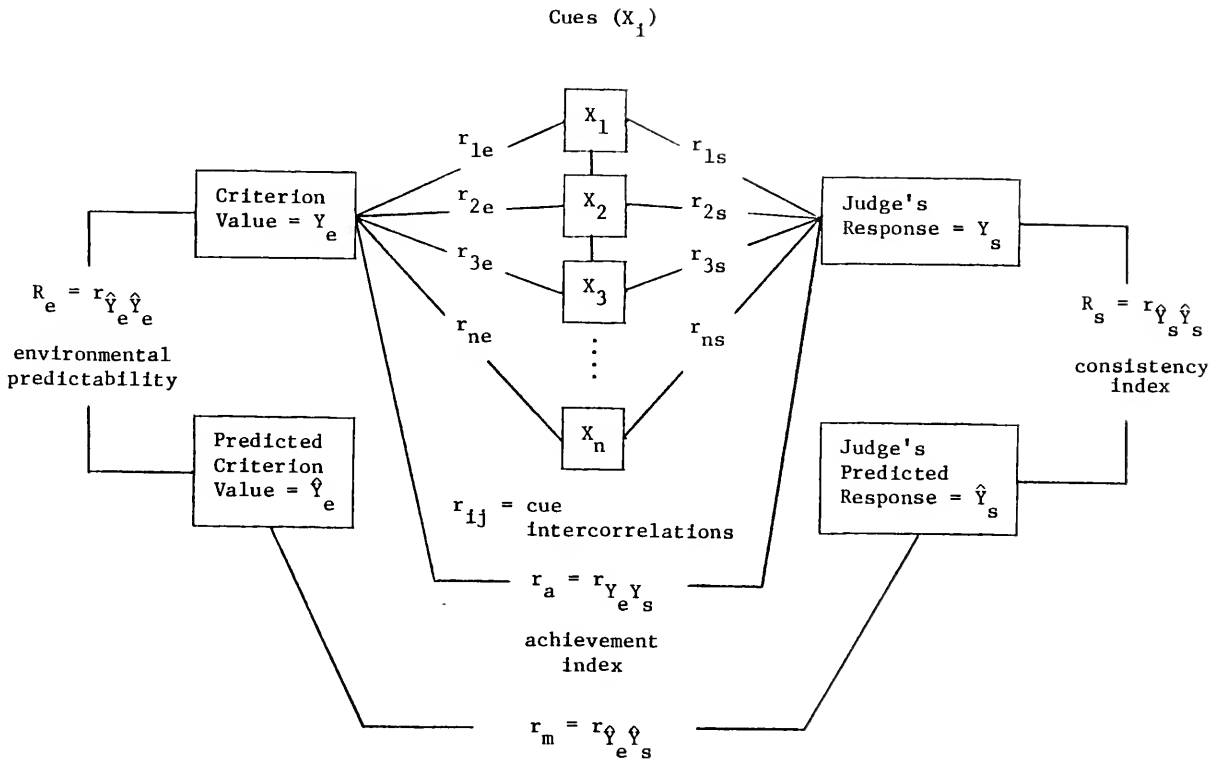
$\hat{Y}_e$ = the predicted criterion value. This prediction results from taking the "optimal" linear combination of the cues, as accomplished by minimizing the squared deviations between Y_e and $\hat{Y}_e$.

Therefore, $\hat{Y}_e = \sum_{i=1}^{k} b_{ie} X_i$,

where b_{ie} represents the relative weightings of the cues.

Y_s = The subject's response. This value represents his judgment as to the criterion value.

$\hat{Y}_s$ = the predicted response of the subject. This prediction results from taking the "optimal" linear combination of



SOURCE: Messier, W. F., Jr., "An Examination of Expert Judgment in the Materiality/Disclosure Decision," Ph.D. dissertation, Indiana University, 1979.

Figure 3-1. Diagram of the lens model showing the relationships among the cues, criteria, and judge's responses.

the cues, as accomplished by minimizing the squared deviations between Y_s and $\hat{Y}_s$.

Therefore, $\hat{Y}_s = \sum_{i=1}^{k} b_{is} X_i$,

where b_{is} represents the relative importance of the cues to the subject.

The "correlational" nature of the lens model is reflected by the six correlations that can be calculated among both the actual and predicted environmental and subject response variables. Within the environmental system (left side) of the model, two correlational measures can be calculated between the criterion value (Y_e) and the cues (X_i), and between the criterion value (Y_e) and the predicted criterion value ($\hat{Y}_e$). The correlations are explained as follows:

r_{ie} = the correlation across stimuli between cue X_i and Y_e. This correlation reflects the relevance of the ith information source (cue) in the environment.

$r_{Y_e \hat{Y}_e}$ = the correlation between Y_e and $\hat{Y}_e$. This correlation represents the degree to which the (linearly) weighted combination of cues serves to predict the state of Y_e and is referred to as "environmental predictability."

Two correlational measures can also be calculated within the subject's system (right side of the lens model). The subject's response (Y_s) can be correlated with the cues (X_i) and also with the predicted subject response ($\hat{Y}_s$). Explanations of these relationships are as follows:

r_{is} = the correlation across stimuli between cue X_i and Y_s. This correlation represents the relevance of the ith information source (cue to the subject).

$r_{Y_s \hat{Y}_s}$ = the correlation between Y_s and $\hat{Y}_s$. This correlation represents the extent to which the subject's judgments

can be predicted by a linear combination of cue values and is referred to as "response linearity."

An evaluation of the subject's performance can be obtained through correlations between the left side and right side of the model. These two measures are known as the "achievement index" and the "matching index" and are explained as follows:

$r_{Y_e Y_s}$ = the correlation between Y_e and Y_s. This correlation represents the subject's ability to predict an outcome and is known as an "achievement index."

$r_{\hat{Y}_e \hat{Y}_s}$ = the correlation between $\hat{Y}_e$ and $\hat{Y}_s$. This correlation is between the two regression equation model estimates and is known as a "matching index."

The use of the lens model in the current internal control evaluation study requires the presentation of a set of relevant accounting information cues (X_i) as a basis for the auditor (subject) to exercise his judgment in deciding the extent of substantive testing to perform (Y_s). The actual decision that should be made using the information cues and the appropriate loss function is represented by Y_e. Because an implied loss function is necessary to decide an "optimal" value for Y_e, this value is not known within the context of the current study.

The lack of a true criterion value (Y_e) is not unusual for real-world judgment situations. As a result, the current study concentrates on the right side of the lens model as the framework for analysis. However, the emphasis on the auditor's responses (Y_s) and his predicted responses ($\hat{Y}_s$) does provide the necessary relationships to properly pursue the objectives of the study. It could be argued that by obtaining a certain level of consensus among the auditors in this

study, an appropriate value for Y_e could be inferred for the specific internal control environment that was predicted.

The Multivariate Lens Model

The existence of more than one criterion value within the context of a single judgmental situation would require some adjustment to the univariate model just presented. Within the current study, the auditor must implicitly decide upon the reliability of the internal control system prior to selecting an appropriate sample size for the substantive test. Figure 3-2 presents the lens model representation of a two criterion value situation.

The multivariate lens model allows for the simultaneous evaluation of a judgment process when there is more than one criterion value. An obvious difficulty associated with the use of the univariate model in such a situation has been summarized as follows: "The researcher would be left with a large number of indices and measures which would require some sort of aggregation in order to interpret exactly what was occurring in the entire system" [Castellan, 1972, p. 244].

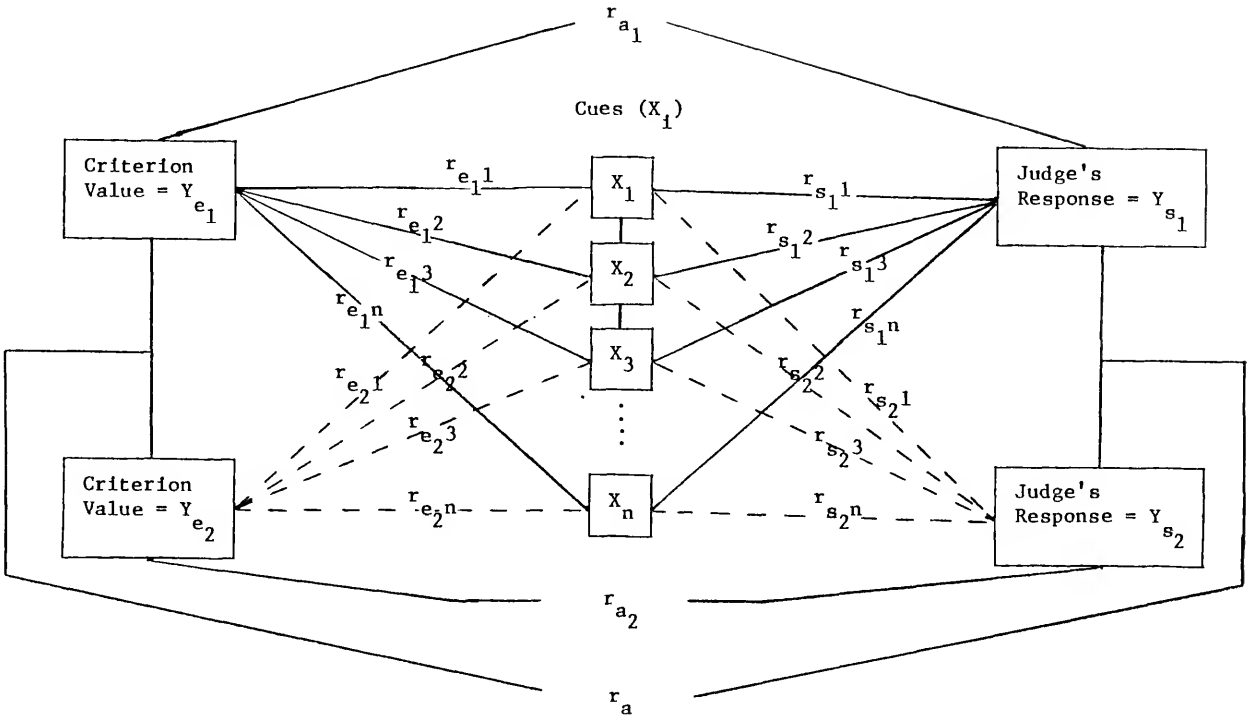
Representation of the two criterion value case requires only minor changes to vector notation as follows:[1]

$\underline{Y}_e$ = the vector (Y_{e1}, Y_{e2}).

$\underline{Y}_s$ = the vector (Y_{s1}, Y_{s2}).

$\underline{X}$ = the vector $(X_1, X_2, \ldots, X_k)$

[1]The following is taken from a more complete discussion of a multiple criterion case of the lens model discussed by Castellan [1972].



SOURCE: Messier, W. F., Jr., "An Examination of Expert Judgment in the Materiality/Disclosure Decision," Ph.D. dissertation, Indiana University, 1979.

Figure 3-2. Diagram of the lens model for the two criterion case showing the relationships among the cues, criteria, and judge's responses.

The multivariate lens model for the two criterion case is then presented as

$$\underline{X}^* = (\underline{Y}_e, \underline{Y}_s, \underline{X}),$$

where $\underline{X}^*$ is a representation of the complete system.

The appropriate correlational measures then become multivariate and can be obtained from the use of canonical correlation. However, the multivariate analysis does not affect the situation where a true criterion value is lacking. Therefore, in the current study, appropriate canonical correlations are found relative to the right side of the lens model.

Analysis of Variance

Strict adherence to the lens model framework discussed above would not allow for the incorporation of a nonlinear model. The analysis of variance (ANOVA) model enables the examination of both linear and configural processing by subjects. The extent to which a cue contributes to a subject's response can be determined through an ANOVA model, as can the significance of the contribution of two or more cues when considered simultaneously. Also, a multivariate analysis of variance (MANOVA) is possible for situations having more than one dependent variable.

The ANOVA (or MANOVA) framework offers some other distinct advantages over a strict adherence to the lens model approach. The use of the factorial design and the orthogonality obtained are specific advantages. The fact that the cues (X_i) can be discrete or categorical rather than continuous offers another advantage for most studies.

And finally, given the desire to examine the auditor's ability to understand his decision process, the calculation of a self-insight index is most important. This is possible through the calculation of omega-squared indices which provide an indication of the proportion of total variance accounted for by a particular cue or combination of cues. Therefore, the framework of the current study is expanded to incorporate the ANOVA model to provide additional important analyses.

The Experimental Design and Methodology

The Task

Task selection. The judgment process used by auditors in routine audit situations is best examined through the use of a specific audit task. The experimental task chosen for this study requires the subject to determine the reliability to be attached to a set of internal controls of a hypothetical company, and then to use the available information about these controls in determining the extent of substantive testing to be performed to meet a specific audit objective (e.g., to satisfy himself that recorded sales are for valid transactions).

As discussed in Chapter II, previous studies dealing with auditor judgment have dealt with a broader audit environment and therefore a somewhat different audit task. The experimental task used in this study was chosen because:

(1) The task is most representative of the type of judgments that auditors must make at different points within the entire audit process.

(2) It incorporates an area (internal control evaluation) that has become a most important topic within the auditing profession.[2]

(3) It represents a very important portion of the audit process because of the commitment of the firm's resources and the risks associated with incorrect decisions.

(4) The task can be represented to subjects in a setting that allows for the best combination of a realistic situation with the necessary control for variables not of specific interest to the experimenter.

(5) It deals with a specific objective within a particular cycle of the business. (A cycle approach to auditing is considered an acceptable and desirable method within the profession.)

Task development. This study focuses on auditors' decisions concerning the reliability of an IAC system and sample sizes for compliance and substantive tests. These decisions are made with respect to a specific audit objective (e.g., satisfaction that recorded sales are for valid transactions) and three key IACs that can be summarized as follows: (1) proper approval for credit, (2) authorization for shipment, and (3) control over physical shipment of goods. Since each of these internal controls could or could not exist, there are eight (2×2×2) possible combinations that could be presented. However, for practical reasons, a number of these "scenarios" were not included in the experiment.

First, since the study is concerned with the effect of compliance test results, the situation where none of the three controls exist was not considered appropriate. This situation also appears to be "unrealistic" in terms of the absence of all three of the key controls.

---

[2]See, for example, the AICPA report by the Commission on Auditor's Responsibility [1978] and The Foreign Corrupt Practices Act of 1977.

The objective of providing settings that would be acceptable as realistic was a major consideration in the selection of scenarios. For example, most companies have procedures to evaluate credit, and it is certainly realistic for a company to have either control over shipment authorization or control over physical shipment without necessarily having both controls. Another consideration in reducing the eight possible scenarios was the amount of time that would have been required to include all of them in the experiment.

As a result of considering the above factors,[3] it was decided that the inclusion of two scenarios would be the most appropriate. Therefore, Scenario #1 includes the existence of all three key internal controls and Scenario #2 includes IC-1 (credit approval) and IC-3 (control over physical shipment). These scenarios provide the necessary information and elicit the appropriate responses for the preliminary audit program stage.

Within each scenario, there are a given number of "cases." These cases provide the opportunity to present compliance test results and elicit responses at the audit program revision stage. The cases are the possible combinations of compliance test results that could occur within a given scenario assuming that a compliance test result will either confirm or disconfirm the auditor's prior belief about the effectiveness of the control. For example, in Scenario #1 where all three of the controls exist, there are eight case situations (2×2×2). In Scenario #2, where two of the three controls exist, there

---

[3]The factors were considered in conjunction with references to auditing texts and audit manuals, along with discussions with faculty members and auditors involved in pilot testing.

are four case situations (2×1×2). Given that no compliance tests are performed if a control is not present, and that possible results of compliance testing will be either confirming or disconfirming, the experimental task at the audit program revision stage consists of twelve case situations.

Decision process. The decision process followed by the auditor in this study is summarized in flowchart format in Figure 3-3 and is discussed in this section. The experimental task requires four responses by the auditor at various stages of the audit process. First, the auditor is asked the degree of reliability he associates with the existing internal controls. This response is made after reviewing information that is common to all case situations, i.e., background information of the company, flowchart of the sales cycle, specific audit concern, etc., and after being presented with the combination of key controls that exist for a particular scenario, e.g., all three controls exist. This first elicitation of reliability (R1) is required for each of the two scenarios and can be represented as follows:

$$R1 = f(B1, I1, I2, I3),$$

where Ri = a measure of reliability taken from a 7-point scale;

B1 = unchanging information, common to all cases;

I1 = internal control #1;

I2 = internal control #2; and

I3 - internal control #3.

Recall that I1, I2, and I3 are dichotomous variables where either: (i) the control exists or (ii) the control does not exist.

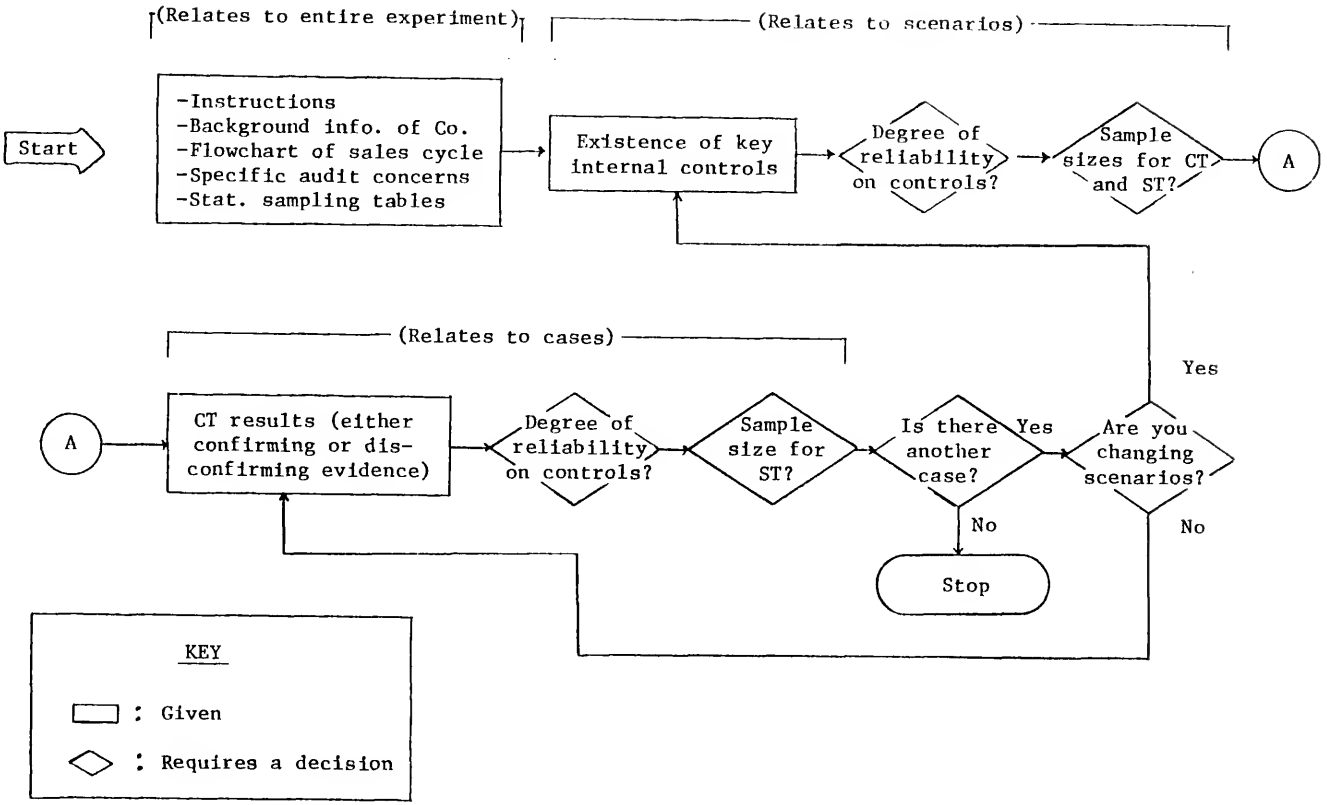


Figure 3-3. The experimental decision process.

At this point, the auditor also is asked for a preliminary audit program[4] (containing planned sample sizes for both compliance tests and a substantive test[5] relating to the specific objective). After receiving specific results in the form of noncompliance rates for the compliance tests, the auditor is asked to indicate the degree of reliability he associates with the existing internal controls. This second elicitation of reliability (R2) is made for all 12 cases and can be represented as follows:

$$R2 = f(BI, I1, I2, I3, R1, C1, C2, C3, PP),$$

where R2 = a measure of reliability taken from a 7-point scale (after observing results of compliance tests);

C1 = result of compliance test for I1;

C2 = result of compliance test for I2, if performed;

C3 = result of compliance test for I3; and

PP = preliminary program as planned by auditor prior to receiving results of compliance testing.

C1, C2, and C3 are trichotomous variables where either: (i) a compliance test was not performed because the control does not exist, (ii) the compliance test resulted in confirming evidence, or (iii) the compliance test resulted in disconfirming evidence.

---

[4] In the context of this study, "preliminary audit program" implies an initial audit program that will possibly be adjusted for the substantive test as a result of the evaluation of compliance test results. At that time a "final audit program" is determined.

[5] The confirmation of accounts receivable was chosen for use in this study because of its most direct relationship with the sales account and its familiarity to the subjects.

The fourth elicitation, a final sample size for the substantive test (FS), also is required for all 12 cases and can be represented as follows:

FS = f(BI, I1, I2, I3, R1, C1, C2, C3, PP, R2).

Experimental materials. The materials provided to the subjects are presented in Appendix A and included the following:

(1) a set of instructions to explain the purpose of the study, the setting in which the subject would operate, the factors that would be varied, and the specific decisions that were to be required of each subject;

(2) background information to describe the company to be audited and additional information that was gathered during preliminary audit work (i.e., the existence of the key internal controls and other controls that may impact on the attainment of the specific objective);

(3) a flowchart representation of the sales cycle of the hypothetical company, with representations of the three key internal controls;

(4) a statement of the specific audit objective to be considered and a summary description of the controls of interest and the corresponding preliminary audit program (containing procedures for both compliance testing and the appropriate substantive test);

(5) specific information relating to each of the 12 individual cases (this includes, for example, the extent to which the three key internal controls exist and the results from performing the desired amount of compliance testing);

(6) tables for determining the sample sizes for both compliance tests and the substantive test using a statistical sampling approach, along with tables for interpreting the results of compliance tests which were presented in the form of noncompliance rates for specific attributes; and

(7) a post-experiment questionnaire.

Cue Selection

The decision to present an audit situation with a specific objective was based primarily on the desire to maximize control over cues provided to the subjects. The audit environment is recognized as very complex with many interrelated components. To say that the three key controls highlighted in this experiment, in conjunction with the accounts receivable confirmation results, are the only important cues in deciding whether "recorded sales are for valid transactions" is of course not true. However, as a result of reviewing auditing textbooks, audit manuals from "Big-Eight" accounting firms, and discussing this particular audit situation with those auditors participating in the pilot studies, the three key controls were considered most appropriate in deciding the number of accounts receivable confirmations to send relative to this particular objective.

Other information that could be deemed important in meeting the specific objective was provided. It was indicated to subjects that the review procedures revealed no problems relating to monthly statements, customer complaints, and prenumbering documents. It also was stated that normal review procedures would be followed in such related areas as inventory cut-off, review of aged receivables, uncollectibles, etc. Appendix A includes the information given the participants concerning these areas and the primary cues to be used in determining the reliability of the system and the resulting extent of substantive testing for this particular audit objective.

Experimental Design

The experiment was conducted using a 2×3×2 factorial design (fixed effects model)[6] and was performed by 109 subjects.[7] A factorial design offers advantages in the evaluation of the cues (factors) used in the experiment. Specifically, factorial designs allow for the evaluation of a combined effect of two or more variables (factors) when used simultaneously. The more complete information that results has been described by Winer:

> Information obtained from factorial experiments is more complete than that obtained from a series of single factor experiments, in the sense that factorial experiments permit the evaluation of interaction effects. An interaction effect is an effect attributable to the combination of variables above and beyond that which can be predicted from the variables considered singly [1971, p. 309].

The impact of interaction effects will vary from study to study. However, it is unusual for any higher order interactions greater than those of two factors to contribute significantly to the model (e.g., Goldberg, 1968). Accordingly, the focus of this experiment upon main effects and two-factor interactions should not constitute a serious limitation.

As discussed previously, two broad scenarios were presented to the subjects. Scenario #1 indicates that all three controls (factors) exist and each control has dichotomous values, either: 1) a

---

[6]See Winer [1971, p. 312] for a discussion of the fixed effects (factors) model.

[7]There were a total of 119 packages distributed. Ten subjects were not included in the analysis because they submitted incomplete packages or responses that indicated a lack of understanding of the requirements of the study.

2%[8] noncompliance rate that implies confirming evidence, or 2) an 8%[9] noncompliance rate that implies disconfirming evidence.

Scenario #2 presents the situation where IC-2 does not exist and therefore will have no compliance testing performed. This allows for a third value (level) for IC-2 compared to only two values for IC-1 and IC-3. Therefore, the experiment has three factors (IC-1, IC-2, and IC-3) with two possible levels for IC-1 and IC-3 and three possible levels for IC-2. This results in the 2×3×2 factorial design with the three factors and their levels as shown in Table 3-1.

The design allows for the estimation of all main effects and two factor interactions with the 11 degrees of freedom available. The degrees of freedom for individual analysis are used as follows:

effect	d.f.
main	4
two factor interaction	5
higher-order (error)	2
total	11

As indicated, the mean square from the higher order interactions was combined to form the error term.

The 12 cases, representing the 12 cells of the factorial design, were presented in a random order within each scenario, with Scenario #1 always presented first. Approximately one-third of the

---

[8]This percentage was given as the "expected" noncompliance rate.

[9]This percentage was developed through discussions with auditors and references to other audit situations. It provides an achieved precision level that is obviously unacceptable and therefore represents "unexpected" or "disconfirming" evidence.

TABLE 3-1

FACTORS AND LEVELS USED IN THE STUDY

Factors	Level	Results of Compliance Tests
1. IC-1: Credit Approval	1	2% noncompliance rate
	2	8% noncompliance rate
2. IC-2: Shipment Authoritation	1	2% compliance rate
	2	8% noncompliance rate
	3	Not performed
3. IC-3: Physical Shipment	1	2% noncompliance rate
	2	8% noncompliance rate

subjects reviewed the cases in one randomized order and the remaining two-thirds in a different randomized order.[10]

Description of the Auditor's Judgment Process

Univariate analysis of variance (ANOVA) was applied to the reliability and sample size responses for the 12 factor combinations of each subject. Also, ANOVA was applied to the responses by groups according to firm and experience levels of the subjects.

F ratios and levels of significance were computed for each main effect and two-factor interaction. As discussed previously, significant interaction effects are indicative of configural processing and are not predominant in human information processing research findings.

The omega-squared statistic was calculated for both reliability and sample size responses to determine the proportion of variance accounted for by each main effect and two-factor interaction. The omega-squared value indicates the proportion of total variance accounted for by a particular factor.

Judgment Consensus

Of major interest in this study is the degree of consensus or agreement among subjects when responding to the same data. The degree of consensus was evaluated at the audit program revision stage using Pearson product-moment correlation, canonical correlation, and cluster analysis.

---

[10]This imbalance resulted from the unexpectedly large number of participants provided by one of the firms.

Pearson correlation measures were calculated among subjects on their reliability decisions and then on the sample size decisions for the 12 cases. This resulted in two measures of the strength of the linear relationship of the responses between each pair of subjects.

Canonical correlation analysis can consider the reliability and sample size resonses simultaneously, enabling the derivation of a linear combination for each subject such that the correlation between each pair of subjects is maximized. This correlation measure is the canonical correlation developed and discussed for the 5,886 possible combinations of subjects in this study.

The cluster analysis was performed both on the reliability and sample size responses individually and combined. The cluster program used a hierarchical clustering technique [Johnson, 1967] which begins by forming a cluster for each subject in the analysis. Using a Euclidean distance measure, the two closest clusters then are combined into one cluster, the two closest of the new set then are combined into a single cluster, and so on.

Self-Insight

Another area of relative importance in this study is the extent of self-insight possessed by the subjects. That is, does the auditor have a good "feel" for the judgment process he employed in arriving at his decisions? This question specifically addresses the auditor's ability to understand and express the relative weights or importance of the cues used to arrive at his decisions.

To allow for the calculation of a self-insight index, a post-experiment questionnaire asked each subject to allocate 100 points to the three key internal controls so as to reflect the relative importance of each to his decisions. A self-insight index was calculated following the approach employed by Slovic [1969], Ashton [1973], and Messier [1979]. The subjective weights elicited from the subjects were correlated with adjusted objective weights calculated from the omega-squared values (discussed earlier). The omega-squared values were adjusted by normalizing the values to 100 for the main effects. The original omega-squared value for a main effect was increased by the omega-squared values for any interaction terms containing that particular main effect. For example, the omega-squared values for the interaction terms (IC-1)(IC-2) and (IC-1)(IC-3) were added to the omega-squared value for the main effect of IC-1. The adjusted omega-squared values for the three internal controls were summed and then each value was divided by that sum. The results are the adjusted objective weights for the three internal controls which sum to 100 and can be correlated with the subjective weights given by the subjects to provide a self-insight index for each subject.

Lack of Formal Hypotheses

The overall objective of this study as given previously is to describe and analyze the auditor's judgment as reflected in a routine audit situation. It should be stressed that due to the emphasis on the descriptive nature of this study there are no formal hypotheses to be tested. However, previous studies concerning auditor judgment

provide some basis for expectations. Accordingly, expected results of the current study are discussed below.

The extent of auditor consensus relative to some recent judgment studies was discussed in Chapter II. The current study is characterized by a similar audit setting requiring similar decisions. Evidence concerning consensus is provided at the preliminary audit program stage, with more extensive evidence provided at the audit program revision stage. Previous studies have indicated various degrees of consensus among auditors. Due to the nature of the audit setting and the audit task, some lack of consensus is expected in this study. However, the existence of some consensus is more likely to appear as the auditors are grouped according to firm affiliation or experience level.

More specifically, it is anticipated that moving from level 1 (confirming evidence) to level 2 (disconfirming evidence) will cause the degree of reliability of the controls to decrease and the sample size for the substantive test to increase. However, the movement to level 3 (control does not exist) from level 2 for factor 2 will not necessarily have the same effect. The question becomes: Which situation is considered "stronger," to have a control that is ineffective in terms of allowing an unacceptable noncompliance rate or for the control not to exist? The control may not exist because it is not needed. For this reason, it is expected that there will be a lack of consensus on this point.

With regard to the decision models developed from the auditors' responses, it is expected that the main effect variables will account for most of the variance when combined in a linear fashion. Two-factor

interactions are expected to account for some variance, with higher-order interactions being insignificant.

Disagreement among the auditors is expected with respect to the relative importance of the three controls (variables). An a priori subjective ranking of the relative importance of the controls within the specific setting of this study is possible, but was not attempted because of the subjective nature of such a ranking. Whatever the relative importance of the variables attached by the auditors, previous accounting studies have indicated a high degree of self-insight by the participants. A similar degree of self-insight is expected in this study.

Administration of the Experiment

Pilot Studies

Several pilot studies were conducted to provide essential feedback on the experimental materials. Two initial sessions were conducted with individual auditors who went through the experiment, followed by a discussion of their comments and suggestions.

The first participant was a member of a local CPA firm who had previously been with a "Big-Eight" firm and has a specific interest in statistical sampling. His comments dealt mostly with clarifications with respect to the representations on the flowchart and the use of the statistical tables provided for the various decisions.

The second participant was chosen primarily because of his affiliation with one of the firms that would be participating in the primary study. His experience level and training also were comparable

to those participants in the primary study. Some changes as to wording and clarity were made as a result of his comments. However, his major contribution was that he felt members of his firm at his experience level would feel "comfortable" with the audit situation presented by the experimental materials.

The final pilot study was conducted in early August, 1979, with three members of one of the primary study firms. Their major concern was that other audit procedures important in this experimental setting must be accounted for in some manner. Their specific recommendations in this area along with some additional changes in wording were incorporated into the materials.

A final revision in the experimental materials was made as the result of discussion with a "Big-Eight" partner with a strong research background. His comments led to some key changes in the wording of the specific objective, a change in one of the key controls, and other important changes necessary for additional "realism" and clarity. The basic structure and organization of the experimental materials and the response modes, however, remained unchanged from earlier pilot studies.

Primary Study

The primary study was conducted with four public accounting firms in August and September, 1979. The study was conducted in the offices of three of the participating firms and in a conference room of a hotel where a seminar was being conducted by the other firm. The experimenter was in charge and present for the duration of all administrations of the experiment. The same set of instructions were read to all participants with no additional comments made or questions answered

during the approximately 70 minutes of each administration. Although there was no time limit given, participants were told to expect to spend 50-60 minutes going through the experiment and actually spent anywhere from 30 to 75 minutes with a mean of about 55 minutes.

The four firms that participated were all members of the "Big-Eight." Three of the firms provided participants from a single office in the same city. The other firm provided subjects from various offices who were brought together through their participation in a statistical sampling seminar. The experience levels of those participating varied somewhat from firm to firm and are summarized in Table 3-2 according to the experience classifications used in Chapter IV to analyze the results of the study. Also, the number of participants from each firm varied somewhat due to the fact that one criterion for choosing subjects was their availability. The major difference was with respect to Firm #4, where 75 usable responses were obtained as compared to 12, 10, and 12 for Firms 1-3 respectively.

Limitations of the Experiment

There are certain inherent limitations associated with behavioral studies. In addition to the problems with the development and conduction of the experiment itself, other shortcomings arise relative to the analysis of the information generated. Specific limitations of this study are discussed below and include:

1. the selection of subjects;
2. the audit environment;
3. the administration of the experiment;
4. the ability to explain the judgment process; and
5. the generalizability of results.

TABLE 3-2

EXPERIENCE LEVEL OF SUBJECTS BY FIRM

		Experience Level		
		E-1 (1 or 2 yrs)	E-2 (≥ 3 yrs)	Total
Firm	Firm #1	—	12	12
	Firm #2	2	8	10
	Firm #3	11	1	12
	Firm #4	39	34	73
	Total	52	55	107

		Experience Level		
		X-1 (1-3 yrs)	X-2 (≥ 4 yrs)	Total
Firm	Firm #1	3	9	12
	Firm #2	6	4	10
	Firm #3	12	—	12
	Firm #4	64	9	73
	Total	85	22	107

NOTE: Subjects 100-4 and 102-4 could not be classified as to experience and are not included in this table.

Selection of Subjects

Studies which require the use of auditors are restricted by the availability of participants. To assemble a group of auditors together for the sake of performing a task such as that required in this study constitutes a substantial intrusion upon their time and therefore is costly to their firms. For this reason, availability becomes a major consideration in determining the participants from each firm. Also, those responsible for the selection of participants from each firm may want their "best" people to represent them in such a study. On the other hand, the "best" people may be out on assignment while the others participate in the "academic exercise" in the office. Whatever the method of selection of participants, it must be recognized as "non-random" and in some cases possibly biased. Whether the results of such a sample represent the cross-section of senior auditors from "Big-Eight" accounting firms is open for discussion. This limitation is therefore noted with the belief that the impact of such a shortcoming on implications of this study is negligible.

Audit Environment

The experimental materials were developed with the goal of presenting the auditors with a realistic setting in which they would have little difficulty responding to the questions presented. A "self-contained" audit situation was sought to the extent possible in order to better control for extraneous variables. However, this results in the sacrifice of a certain amount of realism due to the dynamic and inter-related nature of the audit process.

The resulting audit situation is admittedly lacking in some "realism" characteristics. It was felt that any further attempts at realism would have introduced confounding factors that serve to reduce the internal validity of the experiment. The problem associated with creating a realistic audit setting is discussed in the following section.

Administration of the Experiment

Although the experiment was conducted in a controlled setting by the experimenter, some problems are still recognized. Whether an auditor will make "meaningful" audit decisions in a "contrived" audit situation with no reward or penalty structure must always be of concern. However, positive feedback was received on the materials and the conduct of the experiment with no apparent motivational problems among those participating.

The situation in which seventy-five participants from one firm performed the study in a single room posed some administrative problems. More participants were given the same order of cases as a result of duplication problems and there was not the total lack of communication with others that was assured with the small groups in the other firms. These shortcomings did not seem to create any significant problems in the current study.

Explaining the Judgment Process

To explain the judgment process is a most admirable research goal. This section serves as a reminder that it was not the purpose of this study to explain the judgment process of auditors within the

context of internal control evaluation. There is no attempt to infer any understanding of the judgment process through the evaluation of any statistical techniques such as ANOVA or MANOVA. However, the results of the application of ANOVA procedures are presented as part of the results of this study in an attempt to describe the judgment process.

Generalizability of Results

A non-randomly chosen group of auditors was presented a specific audit situation and asked to evaluate a hypothetical company and respond to a set of questions regarding the extent of audit work to perform. The participants spent approximately sixty minutes in the experiment with no anticipation of reward or fear of penalty. For these reasons, generalizing from the results of this study to other audit situations for these auditors or to other auditors in the same audit situation is not justified.

The inability to generalize the results of this study does not distract from the specific benefits derived and the implications for future research that will be discussed in Chapter V

Summary

This chapter has introduced the framework in which the current study was conducted The lens model was presented along with ANOVA as most appropriate for the analysis of the judgment process of auditors. The dependence upon correlational measures was also emphasized. Within this described framework, the experimental task was discussed,

both as to support for its selection and to the specific experimental materials.

The selection of a 2×3×2 factorial design was discussed in conjunction with the various possible levels for each factor (cue). The analysis techniques that were used within the described framework and to fulfill the objectives of this study were presented. These included the use of correlation (both univariate and multivariate), ANOVA (or MANOVA), cluster analysis, and "self-insight" indices (as developed from omega-squared values).

Without the statement of formal hypotheses, some expected results of this study were presented. While some of these results were inferred from reference to similar previous studies, others were inferred from an evaluation of the professional literature and the practice of auditing. The procedures followed in both the pilot studies and primary study were presented along with information concerning the participants involved.

Finally, certain limitations of the experiment were presented and discussed. None of the limitations are expected to have a significant impact on the results of the study which are presented in Chapter IV.

CHAPTER IV

RESULTS OF THE STUDY

Introduction

The results of the study are presented and discussed in this chapter. As noted in Chapter III, decisions were required by the auditors at both the preliminary audit program stage and at the audit program revision stage. Due to the structure of the study, the analysis of responses from the audit program revision stage is emphasized. This chapter first will present a discussion of the responses relating to the formulation of a preliminary audit program. From examining these preliminary decisions, some comments are made relating to a comparison of the participating firms and to the relative importance of the three IACs used in the study. Finally, and in much greater detail, an analysis of the responses from the audit program revision stage is presented.

The additional data gathered from the twelve case situations presented at the audit program revision stage allow for a more extensive analysis. First, results of the descriptive analysis of the judgment process of the auditors required in their IAC reliability and substantive test sample size decisions are presented. Second, the extent of judgment consensus is evaluated. Finally, the extent of self-insight exhibited by the auditors is discussed. Although some interpretations of the results in these areas are presented within this chapter, the primary evaluation of results and discussion of implications are provided

in Chapter V. The final sections of the present chapter present summaries of additional data gathered in the post-experiment questionnaire.

Preliminary Audit Program

For the preliminary audit program stage, the auditors in the study responded to questions on: (1) the reliability of the IAC system based only on the knowledge of the existence (nonexistence) of the three key internal controls, (2) the desired sample sizes for the compliance tests to be performed on the existing controls, and (3) the preliminary sample size for the substantive test. The limited amount of data gathered from these questions restricts the analysis both within this stage of the audit process and the comparison with the data gathered at the audit program revision stage. However, some interesting relationships can be established from examining responses from the preliminary audit program stage.

Summaries of responses for the preliminary audit program are presented in Table 4-1 and Table 4-2. Table 4-1 provides the mean responses by firm for the reliability and sample size (compliance and substantive test) decisions. Using the data gathered at this point in the audit process, some specific observations can be made regarding the participants and the three key internal accounting controls used in this study.

Comparison Among Firms

The preliminary responses shown in Table 4-1 provide a basis for considering the responses of Firm #1 and Firm #4 to be similar, while the auditors' resonses from Firm #2 and Firm #3 could also be considered similar to one another. In addition, the preliminary responses to the substantive test sample size questions are the basis

TABLE 4-1

MEAN RESPONSES BY FIRM FOR
THE PRELIMINARY AUDIT PROGRAM

SCENARIO #1 - Existence of IC-1, IC-2, and IC-3

	Firm #1	Firm #2	Firm #3	Firm #4	Combined
Reliability	5.00	5.80	5.00	5.11	5.15
CT: IC-1	77.08	122.50	106.25	80.67	86.93
CT: IC-2	93.75	137.50	129.17	99.67	105.73
CT: IC-3	118.75	145.00	179.17	107.93	120.37
Sub. Test	63.33	84.10	74.67	46.83	55.13

SCENARIO #2 - Existence of IC-1 and IC-3

	Firm #1	Firm #2	Firm #3	Firm #4	Combined
Reliability	4.33	4.30	4.27	4.06	4.14
CT: IC-1	77.08	97.50	143.75	88.85	94.44
CT: IC-3	125.00	140.00	181.25	128.37	134.94
Sub. Test	78.83	130.20	110.08	69.81	80.78

for considering Firm #2 and Firm #3 to be more conservative than Firm #1 and Firm #4..

The preliminary reliability responses for the two scenarios are as expected in that the reliability level decreased as IC-2 was eliminated as an existing control. On the average, when moving to Scenario #2, the reliability level decreased by 1.1 (on the 7-point scale); however, for Firm #2 the reliability level decreased by 1.5. Possible explanations are that either Firm #2 is more conservative than the other firms, or that the auditors from Firm #2 perceived IC-2 to be more important than the auditors from the other firms. Evidence of greater importance being attributed to IC-2 is provided by the subsequent analysis of the responses of the auditors (i.e., see Table 4-2 and Table 4-15). In addition, indications of the conservativeness of Firm #2 are found in the auditors' responses to the preliminary decision as to the sample size for the substantive test. For Scenario #1, auditors from Firm #2 suggested a preliminary sample size for the substantive test of 84.10, as compared to a mean response from the auditors of the other firms of 52.20. Similarly, when responding to the same question for Scenario #2, auditors from Firm #2 had a mean response of 130.20 as compared to a mean response of 75.78 for all other auditors.

The above discussion provides evidence of some differences among the firms at the preliminary audit program stage. While some differences exist among the firms in the selection of sample sizes for the compliance and substantive tests, an indication of agreement as to the relative importance of the three IACs is evident.

Relative Importance of IACs

Table 4-2 summarizes the number of auditors who used each of the various confidence levels in testing the three key IACs. Generally, controls that are felt to be of greater importance are tested at a higher confidence level than those controls of lesser importance. Therefore, IC-3 (control over physical shipment) appears to be the most important control across all firms. Although members of Firm #2 indicated the same selection of confidence levels for IC-1 and IC-3 for Scenario #1, the selections for Scenario #2 indicated slightly higher confidence levels for IC-3. IC-2 (shipment authorization) seems to follow IC-3 in importance, with IC-1 (credit approval) considered the least important.[1] As following sections will indicate, these responses are consistent with the more extensively analyzed responses at the audit program revision stage.

Audit Program Revision

The objectives of this study are pursued primarily through the examination of the responses at the audit program revision stage. The twelve case situations presented to the auditors included the results of compliance testing. These tests assumed the use of the suggested sample sizes that were elicited in the preliminary audit program. Due to the number of cases, the data generated from the reliability and substantive test sample size responses allow for a more extensive evaluation and analysis than that from the preliminary audit program

[1] Any statements concerning the relative importance of these three controls are made only within the context of the current study and should not be considered generalizable.

TABLE 4-2

NUMBER OF AUDITORS USING VARIOUS CONFIDENCE LEVELS FOR COMPLIANCE TESTING

SCENARIO #1 - Compliance Testing of IC-1, IC-2, and IC-3

Confidence Level	Firm #1			Firm #2			Firm #3			Firm #4		
	IC-1	IC-2	IC-3	IC-1	IC-2	IC-3	IC-1	IC-2	IC-3	IC-1	IC-2	IC-3
90 %	9	4	1	1	0	1	5	4	2	33	27	18
95 %	3	8	9	8	9	8	5	6	7	38	45	49
97.5%	0	0	2	0	0	0	1	1	2	2	3	7
99 %	0	0	0	1	1	1	1	1	1	0	0	1
Total	12	12	12	10	10	10	12	12	12	73[a]	75	75

SCENARIO #2 - Compliance Testing of IC-1 and IC-3

Confidence Level	Firm #1		Firm #2		Firm #3		Firm #4	
	IC-1	IC-3	IC-1	IC-3	IC-1	IC-3	IC-1	IC-3
90 %	9	1	4	2	3	1	27	14
95 %	3	9	6	7	5	6	41	52
97.5%	0	2	0	0	1	1	5	6
99 %	0	0	0	1	3	4	1	3
Total	12	12	10	10	12	12	74[a]	75

[a]Some auditors indicated zero sample sizes for the compliance testing of IC-1.

stage. The analysis that follows will include an examination of responses made by the individual auditors as well as a comparison of responses among the firms and the various experience levels.

Description of the Reliability/Sample Size Judgment

The experimental cases were arranged in a factorial design (2×3×2) which permitted analysis-of-variance (ANOVA) computations for each auditor participating in the study. ANOVA was used to evaluate the relative importance of the key internal controls provided for this specific reliabilty/sample size judgment.

The ANOVA analysis was performed individually for both the reliability decisions and the sample size decisions. Therefore, an ANOVA model was constructed for each subject corresponding to their reliability decisions and to their sample size decisions. The omega-squared values for each main effect and two-factor interaction are presented by subject (auditor)[2] in Table 4-3 (reliability judgment) and Table 4-4 (sample size judgment).[3] It will be recalled that the omega-squared value indicates the proportion of total variance accounted for by a particular IAC or combination of controls.

As is indicated in Tables 4-3 and 4-4, tests for significant factors could not be conducted for seven auditors with respect to the

---

[2]The auditors are numbered from 1-1 to 109-4, with the second number representing the firm affiliation of the auditor.

[3]ANOVA was originally conducted using the 3-way interaction to form the error term. With a mean-square-error of zero for 24 and 19 auditors on the reliability and sample size judgments, respectively, a second ANOVA was conducted for these auditors using 2-way and 3-way interactions to form the error term. The omega-squared calculations resulting from this second ANOVA are noted with an asterisk (*) in Tables 4-3 and 4-4, and are included in all further anlyses unless otherwise noted.

TABLE 4-3

SUMMARY OF ANOVA RESULTS - RELIABILITY JUDGMENT

Subj - Firm	FACTORS						Total
	IC-1	IC-2	IC-3	IC-1 × IC-2	IC-1 × IC-3	IC-2 × IC-3	
†1 - 1							
*2 - 1	0.0 (.999)	8.8 (.088)	70.2 (.001)				79.0
*3 - 1	0.0 (.530)	37.0 (.005)	44.2 (.001)				81.2
*4 - 1	0.0 (.999)	38.5 (.001)	54.5 (.001)				93.0
5 - 1	40.0 (.008)	44.0 (.015)	8.0 (.038)	0.0 (.500)	0.0 (.423)	4.0 (.125)	96.0
6 - 1	70.2 (.010)	0.0 (.500)	10.6 (.057)	0.0 (.500)	10.6 (.057)	0.0 (.500)	91.4
7 - 1	0.0 (.423)	4.0 (.250)	80.0 (.012)	0.0 (.500)	0.0 (.423)	4.0 (.250)	88.0
8 - 1	4.5 (.184)	18.2 (.125)	53.0 (.207)	9.1 (.200)	0.0 (.423)	0.0 (.999)	84.8
*9 - 1	0.0 (.999)	8.8 (.088)	71.9 (.001)				80.7
†10 - 1							
11 - 1	0.0 (.423)	0.0 (.500)	93.3 (.006)	0.0 (.500)	0.0 (.423)	0.0 (.500)	93.3
12 - 1	0.0 (.243)	4.0 (.250)	80.0 (.012)	0.0 (.500)	0.0 (.423)	4.0 (.250)	88.0
*13 - 2	0.0 (.999)	32.5 (.001)	59.8 (.001)				92.3
14 - 2	0.0 (.622)	44.7 (.058)	37.9 (.035)	0.0 (.750)	2.9 (.225)	0.0 (.500)	85.5
15 - 2	15.6 (.147)	38.6 (.136)	15.6 (.147)	0.0 (.750)	0.0 (.999)	0.0 (.750)	69.8
16 - 2	1.7 (.184)	54.2 (.020)	35.6 (.015)	0.0 (.500)	1.7 (.184)	0.0 (.500)	93.2
17 - 2	0.0 (.423)	0.0 (.500)	66.7 (.038)	0.0 (.500)	0.0 (.423)	0.0 (.500)	66.7
*18 - 2	9.3 (.001)	83.0 (.001)	3.9 (.007)				96.2
19 - 2	0.0 (.423)	21.1 (.039)	52.6 (.008)	0.0 (.500)	0.0 (.423)	21.1 (.039)	94.8
20 - 2	58.6 (.010)	23.7 (.046)	8.9 (.057)	0.0 (.500)	1.8 (.184)	0.0 (.500)	93.0
21 - 2	26.7 (.012)	28.0 (.023)	40.0 (.008)	1.3 (.250)	0.0 (.423)	0.0 (.500)	96.0
*22 - 2	0.0 (.999)	41.6 (.001)	48.6 (.001)				90.2
23 - 3	1.1 (.368)	0.0 (.500)	70.1 (.044)	0.0 (.750)	0.0 (.999)	0.0 (.750)	71.2

Note: The values shown in this table are omega-squared values (significance levels).

† ANOVA was not possible for these subjects due to the lack of variability in their responses.

* These subjects required the use of 2-way and 3-way interactions to form the error term.

TABLE 4-3 - continued

Subj - Firm	FACTORS IC-1	IC-2	IC-3	IC-1 × IC-2	IC-1 × IC-3	IC-2 × IC-3	Total
24 - 3	2.3 (.272)	39.5 (.077)	34.6 (.046)	0.0 (.999)	0.0 (.667)	7.2 (.250)	83.6
*25 - 3	2.3 (.104)	57.5 (.001)	28.7 (.001)				88.5
26 - 3	37.3 (.015)	23.7 (.046)	20.7 (.027)	7.1 (.125)	1.8 (.184)	2.3 (.250)	92.9
*27 - 3	0.0 (.999)	41.6 (.001)	48.6 (.001)				90.2
28 - 3	1.0 (.375)	17.1 (.219)	54.3 (.053)	0.0 (.700)	0.0 (.742)	0.0 (.875)	72.4
29 - 3	26.4 (.044)	15.6 (.125)	42.0 (.029)	0.0 (.750)	0.0 (.999)	3.5 (.300)	87.5
30 - 3	2.5 (.184)	33.0 (.046)	52.1 (.015)	0.0 (.500)	2.5 (.184)	0.0 (.500)	90.1
†31 - 3							
32 - 3	5.0 (.038)	0.8 (.250)	90.9 (.002)	0.0 (.500)	0.0 (.423)	0.8 (.250)	97.5
33 - 3	4.5 (.225)	32.9 (.107)	34.3 (.056)	0.0 (.500)	4.5 (.225)	0.0 (.750)	76.2
*34 - 3	6.7 (.033)	27.9 (.004)	51.9 (.001)				86.5
*35 - 4	19.0 (.014)	38.0 (.008)	19.0 (.014)				76.0
36 - 4	11.7 (.015)	6.7 (.050)	74.3 (.003)	0.0 (.500)	2.8 (.057)	2.2 (.125)	97.7
37 - 4	17.4 (.027)	23.9 (.039)	31.3 (.015)	7.9 (.100)	1.5 (.184)	11.9 (.071)	93.9
38 - 4	7.4 (.096)	66.7 (.026)	7.4 (.096)	3.7 (.250)	0.0 (.423)	3.7 (.250)	88.9
39 - 4	32.4 (.020)	27.0 (.046)	32.4 (.020)	0.0 (.500)	0.0 (.423)	0.0 (.500)	91.8
40 - 4	2.1 (.184)	8.5 (.125)	70.2 (.010)	0.0 (.500)	2.1 (.184)	8.5 (.125)	91.4
*41 - 4	1.8 (.033)	3.7 (.021)	90.8 (.001)				96.3
42 - 4	0.0 (.423)	15.4 (.071)	76.9 (.008)	0.0 (.500)	0.0 (.423)	0.0 (.500)	92.3
43 - 4	10.8 (.020)	16.2 (.026)	64.9 (.033)	5.4 (.071)	0.0 (.423)	0.0 (.500)	97.3
44 - 4	15.2 (.094)	0.0 (.500)	66.4 (.026)	0.0 (.875)	2.4 (.270)	0.0 (.875)	84.0
45 - 4	5.2 (.100)	23.1 (.050)	51.3 (.012)	2.5 (.250)	0.0 (.423)	10.2 (.100)	92.3
46 - 4	5.5 (.118)	4.7 (.219)	81.1 (.011)	0.0 (.500)	0.0 (.742)	0.0 (.875)	91.3

Note: The values shown in this table are omega-squared values (significance levels).

† ANOVA was not possible for these subjects due to the lack of variability in their responses.

* These subjects required the use of 2-way and 3-way interactions to form the error term.

TABLE 4-3 - continued

Subj - Firm	FACTORS						
	IC-1	IC-2	IC-3	IC-1 × IC-2	IC-1 × IC-3	IC-2 × IC-3	Total
47 - 4	3.7 (.184)	0.0 (.500)	77.8 (.015)	0.0 (.500)	3.7 (.184)	0.0 (.500)	85.2
48 - 4	6.8 (.147)	29.3 (.089)	50.8 (.029)	0.0 (.750)	0.0 (.999)	0.0 (.750)	86.9
49 - 4	8.0 (.038)	8.0 (.071)	74.4 (.004)	4.0 (.125)	0.0 (.423)	1.3 (.250)	96.0
50 - 4	13.3 (.038)	33.3 (.031)	44.5 (.012)	0.0 (.500)	0.0 (.423)	2.2 (.250)	93.3
51 - 4	0.0 (.423)	5.9 (.125)	82.4 (.006)	0.0 (.500)	0.0 (.423)	5.9 (.125)	94.2
52 - 4	6.9 (.038)	0.0 (.500)	82.8 (.003)	0.0 (.500)	6.9 (.038)	0.0 (.500)	96.6
53 - 4	0.0 (.423)	21.0 (.100)	63.2 (.020)	0.0 (.500)	0.0 (.423)	0.0 (.500)	84.2
54 - 4	0.0 (.423)	14.2 (.250)	28.8 (.096)	0.0 (.500)	0.0 (.423)	14.2 (.250)	57.2
55 - 4	44.4 (.020)	33.3 (.050)	0.0 (.423)	11.1 (.125)	0.0 (.423)	0.0 (.500)	88.8
56 - 4	3.1 (.368)	6.3 (.429)	18.7 (.225)	0.0 (.750)	0.0 (.999)	0.0 (.999)	28.1
57 - 4	5.7 (.270)	37.7 (.159)	18.3 (.152)	0.0 (.875)	0.0 (.529)	0.0 (.700)	61.7
58 - 4	35.1 (.035)	32.4 (.071)	9.9 (.102)	0.0 (.500)	2.7 (.225)	3.6 (.300)	83.7
59 - 4	9.5 (.102)	40.0 (.058)	33.9 (.035)	0.0 (.750)	2.6 (.225)	0.0 (.500)	86.0
*60 - 4	1.6 (.228)	17.6 (.041)	56.0 (.001)				75.2
61 - 4	3.7 (.270)	34.6 (.125)	12.0 (.152)	0.0 (.500)	3.7 (.270)	11.5 (.250)	65.5
62 - 4	0.0 (.999)	63.2 (.050)	7.6 (.147)	0.0 (.500)	0.0 (.999)	11.7 (.188)	82.5
†63 - 4							
64 - 4	0.0 (.423)	3.7 (.125)	88.9 (.003)	0.0 (.500)	0.0 (.423)	3.7 (.125)	96.3
65 - 4	52.4 (.010)	25.4 (.039)	7.9 (.057)	0.0 (.500)	7.9 (.057)	0.0 (.500)	93.6
66 - 4	2.7 (.100)	12.0 (.050)	74.7 (.004)	1.3 (.250)	0.0 (.423)	5.3 (.100)	96.0
67 - 4	8.2 (.038)	75.3 (.009)	2.8 (.100)	0.0 (.500)	8.2 (.038)	1.4 (.250)	95.9
68 - 4	0.0 (.423)	73.0 (.006)	10.8 (.020)	0.0 (.500)	0.0 (.423)	13.5 (.031)	97.3
69 - 4	6.2 (.038)	28.9 (.017)	57.8 (.004)	3.1 (.125)	0.0 (.423)	1.0 (.250)	97.0

Note: The values shown in this table are omega-squared values (significance levels).

† ANOVA was not possible for these subjects due to the lack of variability in their responses.

* These subjects required the use of 2-way and 3-way interactions to form the error term.

TABLE 4-3 - continued

Subj - Firm	FACTORS						
	IC-1	IC-2	IC-3	IC-1 × IC-2	IC-1 × IC-3	IC-2 × IC-3	Total
†70 - 4							
71 - 4	7.0 (.225)	23.3 (.188)	25 5 (.102)	0.0 (.500)	7.0 (.225)	0.0 (.750)	62.8
72 - 4	3.5 (.270)	0.0 (.500)	74.1 (.034)	0.0 (.875)	0.0 (.529)	0.0 (.875)	77.6
73 - 4	9.6 (.100)	57.1 (.039)	9.6 (.100)	4.7 (.250)	0.0 (.423)	4.7 (.250)	85.7
74 - 4	8.5 (.057)	20.3 (.050)	55.9 (.010)	0.0 (.500)	1.7 (.184)	6.8 (.125)	93.2
75 - 4	0.0 (.423)	4.1 (.100)	92.8 (.003)	0.0 (.500)	0.0 (.423)	0.0 (.500)	96.9
76 - 4	31.2 (.046)	25.9 (.100)	31.2 (.046)	0.0 (.999)	0.0 (.667)	0.0 (.999)	88.3
*77 - 4	0.0 (.999)	7.1 (.088)	77.3 (.001)				84.4
78 - 4	0.0 (.423)	2.6 (.125)	94.0 (.002)	0.0 (.500)	0.0 (.423)	0.8 (.250)	97.4
79 - 4	15.5 (.096)	7.6 (.250)	46.2 (.038)	7.6 (.250)	0.0 (.423)	0.0 (.500)	76.9
80 - 4	0.0 (.423)	17.8 (.036)	76.7 (.004)	0.0 (.500)	0.0 (.423)	1.4 (.250)	95.9
81 - 4	12.6 (.149)	39.8 (.115)	19.4 (.111)	0.0 (.594)	0.0 (.691)	0.0 (.679)	71.8
82 - 4	39.2 (.012)	29.4 (.031)	23.5 (.020)	1.9 (.250)	0.0 (.423)	0.0 (.500)	94.0
83 - 4	0.0 (.493)	15.8 (.250)	47.6 (.069)	0.0 (.813)	0.0 (.808)	0.0 (.500)	63.4
*84 - 4	3.9 (.104)	29.7 (.008)	47.6 (.001)				81.2
85 - 4	71.3 (.030)	0.0 (.875)	13.7 (.118)	0.0 (.700)	0.0 (.742)	0.0 (.875)	85.0
†86 - 4							
87 - 4	1.0 (.272)	4.6 (.200)	84.2 (.009)	0.0 (.999)	0.0 (.667)	3.1 (.250)	92.9
88 - 4	36.4 (.038)	9.1 (.200)	36.4 (.038)	0.0 (.500)	0.0 (.423)	0.0 (.500)	81.9
89 - 4	3.3 (.057)	16.6 (.025)	70.7 (.003)	2.6 (.125)	0.7 (.184)	3.5 (.100)	97.4
90 - 4	7.0 (.057)	45.1 (.020)	29.6 (.015)	0.0 (.500)	1.4 (.184)	11.3 (.071)	94.4
91 - 4	13.0 (.118)	1.2 (.438)	67.7 (.030)	0.0 (.875)	0.0 (.742)	0.0 (.700)	81.9
92 - 4	0.0 (.742)	0.0 (.875)	4.9 (.375)	0.0 (.875)	0.0 (.742)	0.0 (.700)	4.9

Note: The values shown in this table are omega-squared values (significance levels).

† ANOVA was not possible for these subjects due to the lack of variability in their responses.

* These subjects required the use of 2-way and 3-way interactions to form the error term.

TABLE 4-3 - continued

Subj - Firm	FACTORS						
	IC-1	IC-2	IC-3	IC-1 × IC-2	IC-1 × IC-3	IC-2 × IC-3	Total
93 - 4	12.8 (.057)	20.5 (.071)	53.8 (.015)	0.0 (.500)	2.6 (.184)	0.0 (.500)	89.7
94 - 4	10.9 (.038)	7.3 (.100)	76.4 (.006)	0.0 (.500)	0.0 (.423)	0.0 (.500)	94.6
95 - 4	0.0 (.574)	0.0 (.900)	67.5 (.057)	0.0 (.692)	0.0 (.423)	0.0 (.750)	67.5
96 - 4	16.2 (.094)	20.5 (.140)	16.2 (.094)	3.4 (.350)	2.6 (.270)	17.1 (.159)	76.0
97 - 4	42.0 (.029)	15.6 (.125)	26.4 (.044)	0.0 (.750)	0.0 (.999)	3.5 (.300)	87.5
98 - 4	9.6 (.096)	42.9 (.050)	9.6 (.096)	4.7 (.250)	0.0 (.423)	19.0 (.100)	85.8
99 - 4	10.9 (.038)	7.3 (.100)	76.4 (.006)	0.0 (.500)	0.0 (.423)	0.0 (.500)	94.6
*100 - 4	53.6 (.001)	4.9 (.088)	29.7 (.001)				88.2
101 - 4	44.4 (.020)	0.0 (.500)	22.2 (.038)	0.0 (.500)	22.2 (.038)	0.0 (.500)	88.8
102 - 4	51.3 (.012)	38.5 (.031)	0.0 (.423)	2.5 (.250)	0.0 (.423)	0.0 (.500)	92.3
103 - 4	49.1 (.020)	19.5 (.088)	11.5 (.074)	4.2 (.250)	0.0 (.999)	4.2 (.250)	88.5
104 - 4	1.3 (.270)	0.0 (.500)	80.0 (.012)	6.8 (.184)	1.3 (.270)	0.0 (.700)	89.4
†105 - 4							
106 - 4	4.9 (.057)	7.8 (.071)	82.5 (.004)	0.0 (.500)	1.0 (.184)	0.0 (.500)	96.2
*107 - 4	0.0 (.999)	3.1 (.088)	90.2 (.001)				93.3
108 - 4	19.2 (.189)	0.0 (.700)	32.0 (.138)	0.0 (.875)	0.0 (.580)	0.0 (.750)	51.2
109 - 4	3.4 (.074)	87.3 (.007)	3.4 (.074)	1.2 (.250)	0.0 (.999)	1.2 (.250)	96.5
COMPOSITE	4.5 (.001	8.1 (.001)	32.3 (.001)	0.0 (.493)	0.1 (.038)	0.4 (.003)	45.4

Note: The values shown in this table are omega-squared values (significance levels).

† ANOVA was not possible for these subjects due to the lack of variability in their responses.

* These subjects required the use of 2-way and 3-way interactions to form the error term.

TABLE 4--4

SUMMARY OF ANOVA RESULTS - SAMPLE SIZE JUDGMENT

Subj - Firm	FACTORS						
	IC-1	IC-2	IC-3	IC-1 × IC-2	IC-1 × IC-3	IC-2 × IC-3	Total
1 - 1	0.0 (.423)	0.0 (.500)	63.9 (.038)	0.0 (.500)	0.0 (.423)	0.0 (.500)	63.9
2 - 1	0.0 (.423)	2.5 (.250)	87.7 (.007)	0.0 (.500)	0.0 (.423)	2.5 (.250)	92.7
*3 - 1	0.0 (.999)	40.2 (.001)	53.0 (.001)				93.2
4 - 1	5.7 (.184)	11.8 (.196)	52.2 (.033)	0.0 (.500)	5.7 (.184)	1.7 (.409)	77.1
5 - 1	29.7 (.004)	58.8 (.005)	1.3 (.081)	7.8 (.032)	0.1 (.358)	0.7 (.210)	98.4
6 - 1	65.5 (.025)	4.0 (.316)	3.0 (.240)	4.0 (.316)	3.0 (.240)	0.0 (.500)	79.5
7 - 1	0.0 (.423)	15.5 (.008)	61.4 (.001)	0.0 (.500)	0.0 (.423)	22.4 (.006)	99.3
8 - 1	10.7 (.111)	9.1 (.209)	58.7 (.026)	6.0 (.262)	0.0 (.659)	0.0 (.908)	84.5
*9 - 1	0.0 (.999)	5.2 (.088)	83.5 (.001)				88.7
†10 - 1							
†11 - 1							
12 - 1	0.0 (.423)	2.5 (.250)	87.7 (.007)	0.0 (.500)	0.0 (.423)	2.5 (.250)	92.7
13 - 2	3.0 (.044)	67.8 (.004)	20.3 (.007)	5.9 (.045)	0.0 (.423)	1.3 (.158)	98.3
14 - 2	0.0 (.423)	17.8 (.146)	42.0 (.039)	0.0 (.500)	0.0 (.423)	18.3 (.143)	78.1
15 - 2	0.5 (.309)	93.8 (.012)	0.5 (.309)	0.0 (.734)	0.0 (.535)	0.0 (.734)	94.8
16 - 2	1.7 (.184)	33.9 (.031)	55.9 (.010)	0.0 (.500)	1.7 (.184)	0.0 (.500)	93.2
17 - 2	0.0 (.423)	9.8 (.288)	55.6 (.052)	0.0 (.500)	0.0 (.423)	0.0 (.837)	65.4
18 - 2	13.0 (.116)	35.5 (.097)	13.0 (.116)	3.9 (.345)	5.4 (.201)	3.9 (.345)	74.7
19 - 2	0.0 (.423)	25.5 (.001)	63.1 (.001)	0.0 (.500)	0.0 (.423)	11.2 (.002)	99.8
20 - 2	40.2 (.064)	9.4 (.281)	3.2 (.288)	6.3 (.329)	0.0 (.805)	7.1 (.315)	66.2
21 - 2	32.9 (.007)	23.4 (.020)	32.9 (.007)	2.1 (.155)	3.6 (.057)	2.1 (.155)	97.0
*22 - 2	0.0 (.999)	42.5 (.001)	54.8 (.001)				97.3
23 - 3	8.6 (.146)	7.8 (.250)	59.1 (.031)	0.0 (.553)	0.0 (.416)	1.9 (.404)	77.5

Note: The values shown in this table are omega-squared values (significance levels).

† ANOVA was not possible for these subjects due to the lack of variability in their responses.

* These subjects required the use of 2-way and 3-way interactions to form the error term.

TABLE 4-4 - continued

Subj - Firm	FACTORS						
	IC-1	IC-2	IC-3	IC-1 × IC-2	IC-1 × IC-3	IC-2 × IC-3	Total
24 - 3	1.0 (.300)	26.0 (.071)	43.0 (.024)	0.0 (.929)	0.0 (.808)	20.0 (.089)	90.0
25 - 3	0.5 (.369)	41.2 (.066)	45.4 (.032)	0.0 (.737)	0.0 (.869)	0.0 (.803)	87.1
26 - 3	18.6 (.168)	19.3 (.262)	18.6 (.168)	0.0 (.791)	0.0 (.747)	0.0 (.791)	56.5
27 - 3	0.0 (.423)	63.6 (.008)	26.2 (.010)	0.0 (.500)	0.0 (.423)	6.9 (.068)	96.7
28 - 3	1.2 (.356)	19.2 (.189)	48.5 (.052)	0.0 (.541)	0.0 (.721)	0.0 (.518)	68.9
29 - 3	38.2 (.008)	32.6 (.017)	26.0 (.011)	0.0 (.500)	0.0 (.667)	0.0 (.500)	96.8
30 - 3	14.4 (.023)	25.1 (.027)	52.6 (.007)	1.1 (.279)	0.8 (.208)	1.7 (.229)	95.7
31 - 3	30.6 (.124)	7.7 (.370)	3.3 (.332)	0.0 (.587)	1.1 (.387)	0.0 (.643)	42.7
32 - 3	12.4 (.026)	8.1 (.074)	63.2 (.005)	0.0 (.500)	4.1 (.070)	8.1 (.074)	95.9
33 - 3	0.0 (.480)	46.5 (.153)	15.2 (.185)	0.0 (.943)	0.0 (.629)	0.0 (.880)	61.7
34 - 3	6.3 (.012)	37.6 (.004)	45.1 (.002)	0.0 (.500)	6.3 (.012)	3.9 (.036)	99.2
35 - 4	16.4 (.078)	43.8 (.063)	9.6 (.117)	11.5 (.177)	1.9 (.278)	0.0 (.765)	83.2
36 - 4	30.6 (.006)	3.2 (.093)	63.0 (.003)	0.6 (.284)	0.0 (.742)	0.6 (.269)	98.0
37 - 4	0.0 (.423)	64.5 (.014)	20.4 (.021)	0.0 (.500)	0.0 (.423)	9.7 (.079)	94.6
38 - 4	15.1 (.166)	35.8 (.161)	15.1 (.166)	0.0 (.828)	0.0 (.716)	0.0 (.828)	66.0
*39 - 4	49.8 (.001)	0.0 (.999)	49.8 (.001)				99.6
40 - 4	0.0 (.762)	1.0 (.472)	45.7 (.076)	13.2 (.283)	0.0 (.762)	0.0 (.651)	59.9
41 - 4	0.4 (.369)	0.0 (.500)	88.9 (.014)	0.0 (.735)	0.0 (.860)	0.0 (.735)	89.3
42 - 4	0.0 (.423)	6.3 (.107)	82.3 (.005)	0.0 (.500)	0.0 (.423)	6.3 (.107)	94.9
43 - 4	15.2 (.039)	36.7 (.033)	33.6 (.019)	1.8 (.294)	1.7 (.199)	3.2 (.226)	92.2
44 - 4	43.0 (.025)	4.5 (.258)	16.9 (.060)	4.5 (.258)	16.9 (.020)	0.0 (.500)	85.8
45 - 4	5.3 (.197)	39.4 (.085)	21.3 (.076)	3.5 (.348)	0.0 (.751)	8.1 (.250)	77.6
46 - 4	6.4 (.022)	4.5 (.058)	86.3 (.002)	0.8 (.214)	0.2 (.283)	0.1 (.454)	98.3

Note: The values shown in this table are omega-squared values (significance levels).

† ANOVA was not possible for these subjects due to the lack of variability in their responses.

* These subjects required the use of 2-way and 3-way interactions to form the error term.

TABLE 4-4 - continued

Subj - Firm	FACTORS						
	IC-1	IC-2	IC-3	IC-1 × IC-2	IC-1 × IC-3	IC-2 × IC-3	Total
†47 - 4							
48 - 4	9.9 (.105)	16.1 (.130)	46.2 (.028)	2.8 (.333)	4.2 (.184)	4.0 (.293)	83.2
49 - 4	0.0 (.493)	6.8 (.110)	80.6 (.006)	3.3 (.184)	3.2 (.111)	0.5 (.401)	94.4
50 - 4	0.4 (.368)	28.5 (.075)	59.0 (.020)	0.0 (.750)	0.0 (.999)	0.0 (.500)	87.9
51 - 4	14.2 (.041)	0.0 (.500)	71.3 (.009)	1.7 (.300)	3.4 (.130)	1.7 (.300)	92.3
*52 - 4	3.7 (.033)	0.0 (.999)	90.1 (.001)				93.8
53 - 4	0.0 (.831)	26.3 (.151)	48.8 (.051)	0.0 (.749)	0.0 (.485)	0.0 (.608)	75.1
54 - 4	1.5 (.369)	59.3 (.117)	1.5 (.369)	0.0 (.743)	0.0 (.903)	0.0 (.743)	62.3
55 - 4	0.0 (.732)	40.1 (.109)	17.1 (.117)	9.4 (.271)	7.1 (.201)	0.0 (.878)	73.7
56 - 4	2.6 (.368)	0.0 (.500)	33.3 (.147)	0.0 (.750)	0.0 (.999)	0.0 (.750)	35.9
57 - 4	1.0 (.394)	7.3 (.389)	26.5 (.151)	0.0 (.684)	1.0 (.394)	0.0 (.623)	35.8
58 - 4	14.1 (.068)	36.7 (.056)	26.4 (.040)	0.0 (.629)	0.0 (.866)	11.0 (.149)	88.2
59 - 4	2.5 (.208)	20.1 (.087)	63.5 (.016)	0.0 (.500)	2.5 (.208)	0.0 (.731)	88.6
60 - 4	7.3 (.219)	19.6 (.205)	36.1 (.077)	0.0 (.649)	4.2 (.275)	0.0 (.792)	67.2
61 - 4	1.3 (.337)	62.1 (.066)	7.0 (.183)	0.0 (.500)	1.3 (.337)	0.1 (.495)	71.8
62 - 4	0.0 (.423)	80.0 (.013)	7.6 (.062)	0.0 (.500)	0.0 (.423)	5.8 (.138)	93.4
†63 - 4							
*64 - 4	0.0 (.999)	5.2 (.088)	83.5 (.001)				88.7
65 - 4	25.5 (.122)	8.5 (.338)	25.5 (.122)	0.0 (.658)	0.0 (.860)	0.0 (.658)	59.5
66 - 4	1.1 (.321)	13.0 (.160)	67.1 (.022)	0.0 (.871)	0.0 (.926)	4.6 (.286)	85.8
67 - 4	17.2 (.047)	54.8 (.032)	11.6 (.067)	0.9 (.404)	0.0 (.726)	5.1 (.210)	89.6
68 - 4	0.0 (.423)	66.7 (.031)	12.2 (.073)	0.0 (.500)	0.0 (.423)	8.1 (.174)	87.0
69 - 4	0.8 (.333)	19.5 (.107)	66.1 (.019)	0.0 (.824)	0.0 (.873)	1.0 (.418)	87.4

Note: The values shown in this table are omega-squared values (significance levels).

† ANOVA was not possible for these subjects due to the lack of variability in their responses.

* These subjects required the use of 2-way and 3-way interactions to form the error term.

TABLE 4-4 - continued

Subj - Firm	FACTORS						
	IC-1	IC-2	IC-3	IC-1 × IC-2	IC-1 × IC-3	IC-2 × IC-3	Total
†70 - 4							
71 - 4	6.4 (.233)	24.5 (.179)	6.4 (.233)	0.0 (.500)	6.4 (.233)	15.5 (.234)	59.2
72 - 4	0.0 (.423)	0.0 (.500)	66.7 (.038)	0.0 (.500)	0.0 (.423)	0.0 (.500)	66.7
73 - 4	25.4 (.054)	17.5 (.132)	25.4 (.054)	4.6 (.289)	3.8 (.206)	4.6 (.289)	81.3
74 - 4	7.7 (.005)	24.7 (.003)	59.2 (.001)	0.0 (.500)	7.7 (.005)	0.1 (.265)	99.4
*75 - 4	0.0 (.999)	7.1 (.088)	77.5 (.001)				84.6
76 - 4	30.9 (.068)	14.8 (.202)	30.9 (.068)	0.0 (.851)	1.5 (.336)	0.0 (.851)	78.1
77 - 4	0.0 (.423)	39.1 (.018)	54.5 (.007)	0.0 (.500)	0.0 (.423)	1.9 (.216)	95.5
†78 - 4							
79 - 4	9.1 (.184)	0.0 (.500)	45.5 (.057)	0.0 (.500)	9.1 (.184)	0.0 (.500)	63.7
*80 - 4	0.0 (.999)	25.8 (.038)	42.0 (.006)				67.8
81 - 4	13.3 (.077)	29.1 (.074)	23.4 (.048)	10.0 (.168)	0.0 (.935)	10.3 (.166)	86.1
82 - 4	40.9 (.006)	27.7 (.017)	20.2 (.012)	0.0 (.500)	0.3 (.270)	7.8 (.057)	96.9
83 - 4	8.4 (.159)	37.0 (.096)	32.3 (.058)	0.0 (.744)	0.0 (.966)	1.1 (.445)	78.8
84 - 4	2.8 (.050)	3.5 (.077)	78.4 (.002)	0.6 (.253)	6.5 (.023)	6.2 (.047)	98.0
85 - 4	26.6 (.102)	6.4 (.347)	5.0 (.262)	7.0 (.337)	4.3 (.277)	0.3 (.489)	49.6
†86 - 4							
87 - 4	6.6 (.124)	12.5 (.138)	54.8 (.021)	0.0 (.737)	8.4 (.104)	5.0 (.243)	87.3
88 - 4	41.6 (.001)	12.3 (.007)	36.7 (.001)	3.4 (.025)	5.1 (.009)	0.3 (.181)	99.4
89 - 4	1.5 (.029)	37.1 (.003)	51.5 (.001)	3.0 (.029)	0.0 (.423)	6.4 (.014)	99.5
90 - 4	9.6 (.116)	48.6 (.057)	22.0 (.060)	2.4 (.362)	0.0 (.711)	0.2 (.488)	82.8
91 - 4	26.4 (.020)	0.0 (.500)	65.9 (.008)	0.0 (.500)	1.1 (.222)	0.0 (.500)	93.4
92 - 4	0.2 (.368)	44.7 (.026)	20.1 (.028)	0.0 (.750)	0.0 (.999)	29.1 (.039)	94.1

Note: The values shown in this table are omega-squared values (significance levels).

† ANOVA was not possible for these subjects due to the lack of variability in their responses.

* These subjects required the use of 2-way and 3-way interactions to form the error term.

TABLE 4-4 - continued

Subj - Firm	FACTORS						
	IC-1	IC-2	IC-3	IC-1 × IC-2	IC-1 × IC-3	IC-2 × IC-3	Total
93 - 4	8.1 (.022)	39.5 (.009)	45.1 (.004)	2.3 (.123)	4.6 (.060)	0.0 (.500)	97.6
94 - 4	2.9 (.243)	7.2 (.243)	63.1 (.025)	7.5 (.239)	0.0 (.438)	0.0 (.591)	80.7
95 - 4	0.1 (.417)	0.9 (.461)	69.3 (.035)	0.0 (.516)	0.0 (.442)	0.0 (.559)	70.3
96 - 4	0.0 (.434)	0.0 (.527)	70.7 (.044)	0.0 (.671)	0.0 (.675)	0.0 (.816)	70.7
97 - 4	21.3 (.045)	24.7 (.074)	40.7 (.025)	0.6 (.436)	0.0 (.616)	0.5 (.449)	87.8
98 - 4	22.2 (.096)	33.3 (.125)	0.0 (.423)	11.1 (.250)	0.0 (.423)	0.0 (.500)	66.6
*99 - 4	3.3 (.209)	36.0 (.030)	18.0 (.043)				57.3
*100 - 4	48.8 (.001)	0.0 (.999)	48.8 (.001)				97.6
101 - 4	68.4 (.020)	6.0 (.247)	0.0 (.664)	9.8 (.188)	0.9 (.335)	0.0 (.700)	85.1
102 - 4	22.5 (.001)	52.3 (.001)	0.0 (.423)	24.9 (.002)	0.0 (.423)	0.0 (.500)	99.7
103 - 4	39.5 (.098)	4.2 (.415)	1.7 (.368)	5.0 (.402)	0.0 (.999)	0.0 (.750)	50.4
*104 - 4	0.0 (.999)	58.5 (.008)	7.7 (.104)				66.2
†105 - 4							
106 - 4	5.2 (.155)	3.8 (.289)	76.5 (.016)	0.5 (.455)	0.0 (.900)	0.0 (.534)	86.0
107 - 4	0.0 (.721)	13.6 (.258)	35.4 (.081)	0.0 (.500)	0.0 (.721)	13.6 (.258)	62.6
108 - 4	16.7 (.161)	0.0 (.571)	41.2 (.085)	0.0 (.500)	0.0 (.684)	0.0 (.734)	57.9
109 - 4	10.8 (.272)	6.3 (.422)	18.9 (.216)	0.0 (.999)	0.0 (.667)	0.0 (.925)	36.0
COMPOSITE	1.5 (.001)	3.0 (.001)	6.3 (.001)	0.1 (.267)	0.0 (.487)	0.0 (.423)	10.9

Note: The values shown in this table are omega-squared values (significance levels).

† ANOVA was not possible for these subjects due to the lack of variability in their responses.

* These subjects required the use of 2-way and 3-way interactions to form the error term.

reliability judgments and for eight auditors with respect to the sample size judgments. An examination of the responses of these auditors revealed a lack of variability in their responses. In fourteen of the fifteen cases, the auditor's responses fluctuated between only two numbers. For example, auditor number 10 from Firm #1 responded with either a 4 or 2 for the reliability judgment and either 80 or 107 for the sample size judgment. The following discussion of individual ANOVA results is based on the 102 usable results from the reliability judgment and 101 usable results from the sample size judgments.

Results from the individual ANOVAs (Tables 4-3 and 4-4) indicate that IC-3 (control over physical shipment) accounted for most of the variance in their judgments. Specifically, 65 of 102 auditors for the reliability judgment and 57 of 101 for the sample size judgment indicated greatest reliance on IC-3. Auditors whose responses indicated IC-2 (shipment authorization) accounted for most of the variance numbered 21 and 28 for reliability and sample size, respectively, whereas the result for IC-1 (credit approval) indicated 13 and 10 auditors for reliability and sample size. The number of auditors whose responses indicated the highest proportion of variance was attributed equally to IC-1 and IC-3 were 3 and 6 for the reliability and sample size judgments, respectively. Results also indicated that IC-3 was significant ($\alpha \leq .05$) for 73 auditors on the reliability judgment and for 62 auditors on the sample size judgment. Similarly, IC-2 and IC-1 were significant for 42 and 34 auditors, respectively, on the reliability judgment and for 31 and 27 auditors, respectively, on the sample size judgment.

The fact that IC-3 accounted for most of the variance in the auditors' judgments did not preclude IC-1 and IC-2 from accounting for much of the variance in many of the auditors' judgments. For example, in the reliability and sample size judgments, IC-3 accounted for the least amount of variance for 17 and 19 auditors, respectively. Therefore, there is considerable disagreement among some auditors as to the relative importance of the IACs evaluated in this study. This point can be substantiated with respect to the reliability judgment by examining Table 4-5 where omega-squared values for six of the auditors are presented.

The auditors were selected for Table 4-5 as examples of the variance attributed to each of the IACs and their interaction terms. The results are shown in this manner to indicate the relative importance of IC-1, IC-2, and IC-3 to auditors 6-1, 109-4, and 11-1, respectively, and the relative importance of certain interaction terms to auditors 37-4, 101-4, and 19-2. As indicated, auditor number 6-1 relied heavily on IC-1 and to some extent on IC-3 but to no significant extent on IC-2. In sharp contrast, auditors 109-4 and 11-1 relied almost exclusively on IC-2 and IC-3, respectively, when making the reliability judgment.

Referring again to Table 4-3, the importance of interaction effects can be seen. There were 15 auditors who had at least one significant ($\alpha \leq .10$) interaction term in their reliability judgments. Three examples shown in Table 4-5 include auditors number 37-4, 101-4 and 19-2 who reflect significant interactions for IC-1 × IC-2, IC-1 × IC-3, and IC-2 × IC-3, respectively. As can be seen, auditor 37-4 relied on all three IACs and showed two significant interactions,

TABLE 4-5

SELECTED EXAMPLES OF DIFFERENCES IN THE IMPORTANCE OF THE IACs TO THE VARIOUS AUDITORS

(RELIABILITY JUDGMENT)

Auditor Number	Percentage of Total Variance in Judgment Accounted For by IACs:					
	IC-1	IC-2	IC-3	IC-1×IC-2	IC-1×IC-3	IC-2×IC-3
6-1	70.2	ns[a]	10.6	ns	10.6	ns
109-4	3.4	87.3	3.4	ns	ns	ns
11-1	ns	ns	93.3	ns	ns	ns
37-4	17.4	23.9	31.3	7.9	ns	11.9
101-4	44.4	ns	22.2	ns	22.2	ns
19-2	ns	21.1	52.6	ns	ns	21.1

[a] Not significant at the .10 level; all others significant at $\alpha \leq .10$.

while the results for auditor 101-4 indicate no significant reliance on IC-2 and auditor 19-2 indicates none on IC-1 for the reliability judgment. However, both auditors 101-4 (22.2%) and 19-2 (21.1%) have significant interaction terms which account for a large percentage of total variance. Table 4-5, therefore, provides some examples of the variation in responses to the reliability judgments.

Table 4-6 provides a similar analysis for the sample size judgment. A selection of auditors provides results that again indicate the differences that exist with respect to the reliance on the IACs when making the sample size judgment. IC-1 accounted for 68.4% of auditor 101-4's variance, IC-2 accounted for 93.8% of auditor 15-2's variance, and 88.9% of the variance in auditor 41-4's sample size judgment came from IC-3. For all of these auditors, the other IACs and interactions were not significant.

Interaction effects were also significant for many auditors in the sample size judgment. Nineteen auditors had at least one significant ($\alpha \leq .10$) interaction term, with five auditors having two significant interaction terms. As shown in Table 4-6, auditors 88-4 and 32-3 each had two significant interaction terms along with significant main effects for all three IACs. Auditor 92-4 shows a significant interaction for IC-2 $\times$ IC-3, which also accounts for a substantial percentage of his total variance (29.1%). Additional sample items from Tables 4-3 and 4-4 could provide a similar basis for the comparisons discussed above and summarized in Tables 4-5 and 4-6.

The relative importance of the IACs and their two-factor interactions are summarized in Table 4-7. The importance of IC-3 within

TABLE 4-6

SELECTED EXAMPLES OF DIFFERENCES IN THE IMPORTANCE OF THE IACs TO THE VARIOUS AUDITORS

(SAMPLE SIZE JUDGMENT)

Auditor Number	Percentage of Total Variance in Judgment Accounted For by IACs:					
	IC-1	IC-2	IC-3	IC-1 × IC-2	IC-1 × IC-3	IC-2 × IC-3
101-4	68.4	ns[a]	ns	ns	ns	ns
15-2	ns	93.8	ns	ns	ns	ns
41-4	ns	ns	88.9	ns	ns	ns
88-4	41.6	12.3	36.7	3.4	5.1	ns
32-3	12.4	8.1	63.2	ns	4.1	8.1
92-4	ns	44.7	20.1	ns	ns	29.1

[a] Not significant at the .10 level; all others significant at $\alpha \leq .10$.

TABLE 4-7

RELATIVE IMPORTANCE OF THE INTERNAL CONTROLS
AND THEIR TWO-FACTOR INTERACTIONS TO THE SUBJECTS'
RELIABILITY AND SAMPLE SIZE JUDGMENT

Interaction	Number of Subjects For Whom Significant†		Average Proportion of Total Variance Accounted For by Internal Control Variable or Interaction			
			Significant Ones Only†		All Participants	
	Reliability	Sample Size	Reliability	Sample Size	Reliability*	Sample Size**
IC-1	34(49)	28(35)	29.1(22.9)	24.9(25.0)	12.2	11.4
IC-2	42(62)	29(49)	36.7(30.6)	43.5(26.1)	22.4	23.3
IC-3	74(89)	61(78)	57.4(51.3)	52.7(47.9)	46.5	40.7
IC-1 × IC-2	0(2)	5(5)	0(6.7)	9.0(9.0)	1.1	2.0
IC-1 × IC-3	3(6)	4(8)	12.4(9.8)	6.4(6.6)	1.3	1.5
IC-2 × IC-3	2(8)	6(11)	17.3(12.0)	13.2(12.0)	2.8	3.2

† At the .05 level of significance.

()At the .10 level of significance.

* From 102 subjects for which ANOVA analysis was possible.

** From 101 subjects for which ANOVA analysis was possible.

the specific audit setting used in this study is again evident, with evidence for IC-2 being considered more important than IC-1. The relative importance of the controls is maintained whether evaluating the reliability judgment or the sample size judgment and also for either participants with significant controls and interactions or for all participants.

The fact that the least important control (IC-1) accounts for 24.9% of the variance in significant sample size judgments and 11.4% of the variance in all sample size judgments is an indication that none of the three controls was disregarded completely. The dependence, to a certain extent, on all three controls is reflected in the relative importance of the interaction terms. For significant interactions, the average variance was 14.9% for the reliability judgment and 9.5% for the sample size judgment. Comparisons to other studies should be made with caution due to the specific nature of the audit task and the inclusion of IACs known for their importance in making the required judgments. However, other judgment studies in auditing have shown average variances for significant interaction terms of 3.3% [Joyce, 1976], 5% and 5.1% [Messier, 1979], and 6.4% [Ashton, 1974].

An evaluation of the variances for the interaction terms across all participants indicates average variances of 1-3%, which may be considered "normal" for this type of task. An analysis of the information pertaining to the interaction terms indicates that few of the auditors had at least one significant ($\alpha \leq .05$) interaction term (5 of 85 for reliability judgment; 10 of 90 for sample size judgment)[4]

[4]The weightings of the interaction terms were possible for 80 and 95 auditors, respectively, for the reliability and sample size judgments as previously indicated.

although average variances were high for the significant interactions. It therefore appears that some auditors processed information in a highly configural manner, although the number was few.[5]

The various firms and experience levels of the auditors may be suggested as factors underlying differences in responses. Table 4-8 provides a summary of the omega-squared values by firm and experience level. Two classification schemes are used in presenting information as to experience. First, a comparison is made between those auditors with 1-2 years of experience and those with 3 or more years. Secondly, those with 3 years of experience are grouped with those that have 1-2 years and contrasted with the auditors with 4 or more years of experience. The values that are shown in the table are averages that were computed by summing the omega-squared values for a particular firm or experience level and dividing by the number of auditors in that category. The total of 102 auditors for the reliability judgment and 101 auditors for the sample size judgment for which the ANOVA analysis was possible are reduced to 100 and 99, respectively, for the experience classification due to the inability to properly classify two of the auditors.

IC-3 appears to be recognized as the most important control for the reliability decision irrespective of the firm or experience level. However, for Firm #2 the difference between IC-3 and IC-2 is negligible. The relative importance of IC-3 is maintained for the sample size judgment across all classifications except for Firm #2, where the average omega-squared value of IC-2 (35.9) was greater than

---

[5]See, for example, auditors 88-4 and 89-4 in Table 4-4.

TABLE 4-8

VALUE OF STATISTICAL WEIGHTS BY FIRM AND EXPERIENCE LEVEL

Firm or Experience Level	No. of Subjects	Average Statistical Weights for Internal Control Variables: Reliability					
		IC-1	IC-2	IC-3	IC-1 × IC-2	IC-1 × IC-3	IC-2 × IC-3
Firm # 1	(10)	11.5	16.3	56.6	1.5	1.8	2.0
Firm # 2	(10)	11.2	36.7	37.0	0.2	0.9	3.0
Firm # 3	(11)	8.1	26.3	48.0	0.9	1.1	1.7
Firm # 4	(71)	13.1	20.6	46.2	1.2	1.3	2.9
Exp. E-1 (1-2)	(49)	9.9	21.9	47.5	1.2	1.1	3.5
Exp. E-2 (≥3)	(51)	12.8	22.9	46.7	1.1	1.6	2.1
Exp. X-1 (1-3)	(79)	13.1	22.3	44.3	1.0	1.3	3.2
Exp. X-2 (≥4)	(21)	5.3	22.4	57.4	1.4	1.5	1.2
		Sample Size					
		IC-1	IC-2	IC-3	IC-1 × IC-2	IC-1 × IC-3	IC-2 × IC-3
Firm # 1	(10)	11.2	14.9	55 3	2.2	1.1	3.7
Firm # 2	(10)	9.1	35.9	34.1	2.0	1.2	4.9
Firm # 3	(12)	11.0	27.9	37.2	0.1	1.0	3.5
Firm # 4	(69)	11.8	21.9	40.2	2.4	1.7	2.8
Exp. E-1 (1-2)	(50)	9.6	24.9	41.8	1.1	1.7	3.6
Exp. E-2 (≥3)	(49)	12.2	21.5	40.2	2.5	1.3	2.9
Exp. X-1 (1-3)	(79)	11.8	24.9	38.1	1.9	1.6	3.1
Exp. X-2 (≥4)	(20)	7.2	16.5	52 5	1.2	1.5	4.3

IC-3 (34.1). This was due primarily to the omega-squared values of IC-2 (93.8) and IC-3 (0.5) for auditor 15-2. Eliminating his values from the sample size judgments of Firm #2 would result in average values of 29.5 (IC-2) and 37.8 (IC-3) for the remaining nine auditors.

Of additional interest with respect to the judgments made by the auditors is the total variance explained by the individual ANOVA models. By summing the total omega-squared values for the IACs and their interaction terms, an indication of the overall percentage of variance in their responses that is accounted for by the IACs and interactions is obtained. Table 4-9 provides a distribution of auditors by total variance explained.

A large percentage of auditors had omega-squared values of at least 80%. Specifically, 79% of the auditors for the reliability judgment and 61% for the sample size judgment had total omega-squared values of at least 80%. These results may appear misleading when compared to the omega-squared values computed for the composite judgments of all auditors (45.4% for reliability and 10.9% for sample size). These composite results indicate some disagreement among auditors with respect to the reliability decision and substantial disagreement with respect to the sample size decisions.

The findings presented in this section with regard to the description of the judgments made in this study generally are not surprising. That is, there was some disagreement as to the relative importance of the IACs in terms of meeting a specific objective, and this was reflected in decisions concerning reliability and sample size. On the other hand, there was general agreement that IC-3 was of most

TABLE 4-9

DISTRIBUTION OF SUBJECTS BY TOTAL VARIANCE EXPLAINED

Range of Omega-Squared Values	Reliability Judgment	Sample Size Judgment
.90 - 1.00	48	38
.80 - .89	33	24
.70 - .79	10	13
.60 - .69	7	14
.50 - .59	2	7
≤ .50	2[a]	5[b]

[a] 1 = 28.1
2 = 4.9

[b] 1 = 35.9
2 = 35.8
3 = 49.6
4 = 36.0
5 = 42.7

Note: Subjects with omega-squared values of ≤.50 had little variability in their responses.

importance in making the required decisions, followed by IC-2 and IC-1 in that order. As anticipated, most of the variance in the responses of the auditors was captured by a linear combination of the IACs. The interaction terms of the analysis did provide some of the total variance. Indications of configural processing were found primarily in individual auditor responses. And finally, the composite model of the auditors' judgments did a very poor job of explaining the variation in the responses. This was especially true of the sample size judgments.

Judgment Consensus

The extent to which the auditors agreed in their decisions given the same set of data, which is known as judgment consensus, was evaluated through the use of canonical correlation, Pearson product-moment correlation, and cluster analysis. The results of applying these techniques are presented in this section. The results obtained from performing these analyses also will be discussed with respect to firm and experience effects.

The ability to properly evaluate the extent of consensus required an assumption of usable and reliable responses by the auditors. As a measure of the auditors' abilities to understand the task and respond in an internally consistent manner, a Pearson product-moment correlation was computed across the 12 cases between the reliability and sample size responses for each auditor. Given the nature of the decisions, it was hypothesized that as the reliability measure increased (decreased), the suggested sample size would decrease (increase). A negative Pearson product-moment correlation measure was

assumed to indicate "internal" consistency. Due to the scale of the responses for reliability (7-point scale) and sample size (0 to 959), the correlation measures varied from .37 to -1.00, with an average correlation of -.46. The responses of the auditors who had positive correlations (there were a total of 6) were examined for reasonableness. All of the auditors were exhibiting sufficient internal consistency to consider their responses as conscientious efforts, except for an auditor in Firm #1. This auditor's responses suggested a lack of understanding or motivation concerning the task and his responses were excluded in their entirety as indicated in Chapter III.

Canonical correlation. The extent of agreement among auditors with respect to their multivariate decisions on reliability and sample size was evaluated through the analysis of canonical correlation measures. Canonical correlation analysis can consider the reliability and sample size responses simultaneously, enabling the derivation of a linear combination for each subject such that the correlation between each pair of subjects is maximized. Since 109 auditors participated in the study, there are a total of 5,886 canonical correlations between all pairs of subjects.

The mean canonical correlation was .78 with a range of .01 to 1.00. Similar auditor judgment studies that have computed canonical correlations show similar results. For example, Joyce [1976] reported a mean correlation of .93 with a range of .36 to 1.00 and Messier [1979] reported a mean correlation of .75 with a range of .24 to .98. Although these correlations all appear to be rather high, the only conclusion that can be drawn is that the linear combinations of one

particular auditor's judgments seem to be good predictors of linear combinations of another auditor's judgments in these specific situations. An assessment of the variability of the correlations in this study resulted in a standard deviation of .145 which compares favorably to the standard deviation of .17 reported by Messier [1979]. However, it still provides an indication of some variability in the correlations.

Pearson product-moment correlation. Consensus among auditors also can be evaluated by examining the Pearson product-moment correlation between each pair of auditors for both the reliability and sample size judgments. This approach resulted in two measures of the strength of the linear relationship of the responses between each pair of auditors. The mean Pearson product-moment correlation for the reliability judgment was .693 with a range of -.22 to 1.00. The sample size judgment resulted in a mean of .609 with a range of -.66 to 1.00.

The Pearson product-moment correlations found in this study may be considered rather high when compared to Joyce [1976], who reported an average consensus measure of .373 with a range of -.68 to .93. However, there have been similarly high consensus measures obtained in audit settings, such as those reported by Ashton [1974] and Messier [1979]. Ashton obtained a mean correlation of .70 with a range of .06 to .93. Messier, in measuring both a materiality and disclosure judgment, observed mean values of .665 and .670 with ranges of -.10 to .95 and -.05 to .98, respectively.

The variability of the correlation measures, as reflected in their standard deviations, may be expected to be high due to the wide ranges of results indicated above. The standard deviation of the

correlations for the reliability and sample size judgments were .19 and .22, respectively. The finding of relatively high mean correlations with significant variability is consistent with the results of the canonical correlation analysis discussed previously.

Firm and experience effect. To evaluate the possible impact of firm or experience differences on the correlation results, average correlation measures were calculated by firm and experience level. Table 4-10 presents the average canonical correlation by firm and experience level. As indicated previously, two possible experience level categorizations are utilized to determine what possible effect this may have on the analysis. The "E" categories contrast experience levels of 1-2 years with 3 or more years, while "X" categories use 1-3 years versus 4 or more. The values found in Table 4-10 represent the extent of agreement among auditors within the same firm or experience level (diagonal values) or the extent of agreement between two firms or experience levels (off-diagonal values). Results indicate that by combining auditors with 3 years experience with the group with 1-2 years experience as opposed to those with 4 or more years experience, the mean canonical correlation of responses for the more experienced auditors increases from .766 to .790.

Table 4-11 provides data on the effect of firms and experience levels on the reliability and sample size judgments as reflected in Pearson product-moment correlation measures. These results are consistent with the canonical correlations. That is, some minor variations exist between firms and the categorization of experience levels can have an effect on the results. All firms exhibited relatively high

TABLE 4-10

JUDGMENT CONSENSUS AMONG AUDITORS BY FIRM

Firm	(1)	(2)	(3)	(4)
(1)	.770	.785	.797	.776
(2)		.826	.830	.778
(3)			.827	.781
(4)				.774

JUDGMENT CONSENSUS AMONG AUDITORS BY EXPERIENCE LEVEL

Experience Level		(E-1)	(E-2)
(E-1)	1-2 years	.796	.783
(E-2)	3 or more		.766

JUDGMENT CONSENSUS AMONG AUDITORS BY EXPERIENCE LEVEL

Experience Level		(X-1)	(X-2)
(X-1)	1-3 years	.785	.783
(X-2)	4 or more		.790

Note: The values in this table are mean canonical correlations.

TABLE 4-11

JUDGMENT CONSENSUS AMONG AUDITORS BY FIRM
- RELIABILITY (SAMPLE SIZE JUDGMENT) -

Firm	(1)	(2)	(3)	(4)
(1)	.692(.612)	.678(.587)	.710(.629)	.688(.603)
(2)		.721(.637)	.739(.679)	.688(.603)
(3)			.751(.708)	.696(.612)
(4)				.688(.603)

JUDGMENT CONSENSUS AMONG AUDITORS BY EXPERIENCE LEVEL
- RELIABILITY JUDGMENT -

Experience Level		(E-1)	(E-2)
(E-1)	1-2 years	.715	.693
(E-2)	3 or more		.685

Experience Level		(X-1)	(X-2)
(X-1)	1-3 years	.698	.693
(X-2)	4 or more		.734

- SAMPLE SIZE JUDGMENT -

Experience Level		(E-1)	(E-2)
(E-1)	1-2 years	.651	.609
(E-2)	3 or more		.580

Experience Level		(X-1)	(X-2)
(X-1)	1-3 years	.619	.609
(X-2)	4 or more		.637

Note: The values in this table are mean Pearson correlations.

Pearson product-moment correlations both within each firm (diagonal values) and between each combination of two firms (off-diagonal values). Firm #3 had the highest intra-firm correlation ($r = .751$) and had inter-firm correlations with the other firms that were higher than the intra-firm correlations of any other firm.

Of additional interest in Table 4-11 are the correlations for the various experience levels. The highest correlations are obtained in the 1-2 years (E-1) classification and the 4 or more (X-2) classification. If those auditors with 3 years experience are grouped with those that have 1-2 years experience, the correlation decreases from .715 (.651) to .698 (.619) for the reliability (sample size) judgment. If the auditors with 3 years of experience are grouped with those that have 4 or more years experience, the correlation decreases from .734 (.637) to .685 (.580) for the reliability (sample size) judgment.

In an attempt to further analyze the significance of firm or experience levels on the judgments made by the auditors, an ANOVA analysis was performed.[6] The ANOVA analysis combined the firm factor with the three IACs for the evaluation of the firm significance and the experience factor (both E and X classifications) with the three IACs for the evaluation of the significance of experience. For both the reliability and sample size judgments, the firm was considered as having a significant ($\alpha \leq .01$) effect. Using the "E" classification scheme (1-2 years versus 3 or more) the experience level effect was

---

[6] A combined MANOVA was performed that indicated the firm affiliation and the experience level were significant ($\alpha \leq .01$). MANOVA results also provided evidence that the three IACs were significant ($\alpha \leq .01$), as were two of the interaction terms: IC-1 × IC-3 ($\alpha \leq .05$) and IC-2 × IC-3 ($\alpha \leq .01$).

considered significant for both the reliability judgment ($\alpha \leq .05$) and the sample size judgment ($\alpha \leq .01$). In contrast, using the "X" classification scheme (1-3 years versus 4 or more) the experience level was considered significant ($\alpha \leq .05$) only for the reliability judgment.

An additional analysis of the significance of differences in responses as a result of firm and experience differences is shown in Table 4-12. These results are similar to those found at the preliminary audit program stage and presented in the first section of this chapter. These results also indicate the apparent "conservatism" exhibited by Firm #2 in their sample size decisions. That is, their sample size selections are consistently much higher than those of the other firms, and their reliability measures are comparable if not higher in some cases.[7] Also, the responses from Firm #4 are comparable to the other firms with respect to the reliability judgment, yet the auditors indicate on the average much smaller sample sizes. The responses of Firm #4 at the audit program revision stage, therefore, provide an interesting contrast to those of Firm #2.

An evaluation of the average responses shown in Table 4-12 according to the various experience classifications also yields some interesting results. An examination of the "E" classifications suggests that the more experienced auditors were less conservative in choosing a reliability level (e.g., consistently greater reliability across all levels of controls), yet consistently chose higher sample sizes. Results from examining the "X" classifications show mixed

---

[7] It is important to remember that greater reliability is expected to result in smaller sample sizes, not larger ones.

TABLE 4-12

MEAN RESPONSES BY FIRM AND EXPERIENCE LEVEL

RELIABILITY

Variable	Level	Overall	Firm #1	Firm #2	Firm #3	Firm #4	(1-2) E-1	(≥3) E-2	(1-3) X-1	(≥4) X-2
IC-1	1	4.08	3.63	4.22	4.28	4.10	4.03	4.12	4.08	4.06
	2	3.43	3 18	3.65	3.63	3.41	3.37	3.49	3.38	3.61
IC-2	1	4.39	3.94	4.85	4.58	4.38	4.32	4.46	4.34	4.59
	2	3.42	3.04	3.65	3.81	3.39	3.39	3.46	3.40	3.52
	3	3.44	3.23	3.30	3.46	3.49	3.39	3.48	3.46	3.38
IC-3	1	4.63	4.36	4.55	4.74	4.67	4.58	4.68	4.59	4.79
	2	2.88	2.44	3.32	3.17	2.84	2 81	2.93	2.87	2.88

SAMPLE SIZE

Variable	Level	Overall	Firm #1	Firm #2	Firm #3	Firm #4	(1-2) E-1	(≥3) E-2	(1-3) X-1	(≥4) X-2
IC-1	1	84.81	106.57	144.28	110.24	68.90	76.08	93.06	82.42	93.52
	2	113.36	114.08	218.57	141.46	94.21	96.85	128.97	113.47	112.97
IC-2	1	71.53	99.25	112.58	87.79	58.68	62.90	79.70	69.15	80.26
	2	108.26	116.31	207.63	114.52	92.29	90.40	125.14	105.86	117.02
	3	117.46	115.42	224.08	175.23	93.69	106.09	128.20	118.82	112.46
IC-3	1	70.66	81.78	137.08	92.54	56.13	61.27	79.53	69.33	75.51
	2	127.51	138.88	225.77	159.15	106.98	111.66	142.49	126.56	130.99

responses when comparing the more experienced auditors to the less experienced.

An important question can be raised with respect to the responses at levels 2 and 3 of IC-2. That is, how did the auditors perceive the distinction between a specific control not being present (level 3) versus evidence in the form of compliance test results that indicate a control is not operating effectively (e.g., results show a very high noncompliance rate) (level 2)? As could probably be expected, the average responses for levels 2 and 3 of IC-2 show mixed reactions to these situations. Overall, results indicate auditors feel the reliabilities of the IAC systems are comparable, although slightly greater sample sizes would be taken if the control were not present as compared to the control not operating effectively. In contrasting individual firms, note that Firm #4 provides responses for the reliability judgment that indicate a slightly greater reliability for level 3 as compared to level 2 for IC-2, whereas the other firms' responses indicate that more of a difference does exist. However, the responses of Firm #2 and Firm #3 suggest a perception that level 2 represents a more reliable system than level 3; the auditors in Firm #1 indicated just the opposite. The sample size responses for Firm #1 (116.31; 115.42) and Firm #4 (92.29; 93.69) are very similar for the two levels. However, both Firm #2 (207.63; 224.08) and Firm #3 (114.52; 175.23) show average responses that indicate a much larger sample size is needed in the situation where IC-2 is not present than when it is considered not to be operating effectively.

Cluster analysis. The cluster analysis was performed to determine to what extent the auditors who were grouped together in the analysis had any similar characteristics. The data used for clustering the auditors into groups were the standardized[8] responses to the reliability and sample size decisions for the twelve cases. Cluster analysis was run on the reliability and the sample size responses individually, and also on the judgments taken together.

The results of the cluster analysis provide for another measure of the degree of consensus. The results indicate that the groupings for the reliability judgment bear no relationship to those for the sample size judgment. This result is not surprising given the large scaling difference between the two responses. Furthermore, the standardization of the responses could not keep the sample size decisions from dominating in the combined cluster analysis. In other words, the clusters that resulted from using the standardized reliability and sample size decisions together were the same as those that resulted from using the sample size decisions only.

A summary of the cluster analysis is presented in Table 4-13. The results for the movement from four clusters to two clusters indicate that there are some outliers (18, 20, 27, 81) for the sample size and combined analysis. A classification of the auditors in each cluster as to firm affiliation and experience level could not explain the make-up of the clusters. No specific conclusion can be stated concerning consensus from the results of the cluster analysis.

---

[8]Standardization was accomplished by taking the mean response for each auditor across the twelve cases (for both the reliability and sample size judgments) and subtracting each of his responses.

TABLE 4-13

SUMMARY OF CLUSTER ANALYSIS

- SAMPLE SIZE AND RELIABILITY RESPONSES -

	Cluster	Auditors
No. of clusters = 4	1.	1, 2, 7, 11, 15, 16, 17, 26, 29, 35, 46, 48, 49, 55, 61, 63, 74, 85, 95
	2.	All others
	3.	18, 20, 27
	4.	81
No. of clusters = 2	1.	All others
	2.	18, 20, 27, 81

- RELIABILITY RESPONSES -

	Cluster	Auditors
No. of clusters = 4	1.	All others
	2.	1, 3, 8
	3.	5, 18, 20, 26, 35, 55, 61, 65, 67, 68, 76, 82, 97, 102, 103
	4.	32, 36, 41, 46, 52, 64, 75, 78, 87, 105, 107
No. of clusters = 2	1.	All others
	2.	32, 36, 41, 46, 52, 64, 75, 78, 87, 105, 107

Self-Insight

Self-insight indices were computed for each auditor to examine the extent of agreement between the auditor's subjective evaluation of the relative importance of the three key IACs in making his reliability/sample size decisions and their importance as reflected in the statistical weights determined through the ANOVA results.[9] The auditors' subjective weights were elicited directly in the post-experiment questionnaire by asking each auditor to allocate 100 points among the three IACs according to their relative importance to the reliability/sample size decisions in the 12 cases just completed. The statistical weights used in arriving at the self-insight indices were the adjusted omega-squared values for the three key IACs; these reflect the proportion of total variance in the auditor's judgment accounted for by each control.

The use of the omega-squared values for both the reliability and sample size judgments resulted in two sets of statistical weights for each auditor and, therefore, enabled self-insight indices for the reliability judgment and for the sample size judgment to be obtained. These indices are discussed in greater detail in the following sections. First, the subjective weights given by the auditors are discussed. Then, the results of the computation of the self-insight indices are presented and discussed.

---

[9]Self-insight indices could be calculated for only those auditors who responded to the appropriate questions for subjective weights and those for which omega-squared values were available from the ANOVA analysis. Three subjects failed to indicate subjective weights in their post-experiment questionnaire.

Subjective weights. The weights assigned to each IAC by the auditors are presented in Table 4-14. A comparison of the overall average responses for each factor indicates that IC-3, with an average weight of 48.1, is considered the most important. IC-2 (28.9) ranked a distant second, although thought to be slightly more important than IC-1 (23.0). A comparison of these averages with those shown in Table 4-7 indicates that as a group the auditors were able to assess the relative importance of their most important cue very successfully, although perhaps slightly overstating its importance. The importance of IC-2 to the auditors' judgments also was reasonably well recognized by the auditors themselves. However, consistent with similar previous research dealing with human judgment [Slovic and Lichtenstein, 1971], the auditors tended to overestimate the relative importance of the least important cues. In this study, though IC-1 was considered by the auditors to be least important, an average subjective weight of 23.0 was provided by them; on the other hand, the ANOVA results indicated average statistical weights of 12.2 and 11.4 for the reliability and sample size judgments, respectively.

A summary of the subjective weights by firm and experience level is provided in Table 4-15. A comparison of firms indicates that members of Firm #1 and Firm #4 weighted the importance of IC-3 slightly higher than did the other firms, and correspondingly weighted IC-2 slightly lower. The lowest rating of IC-1 was provided by Firm #1; Firm #2 weighted IC-1 lower than Firm #4 due to the greater subjective weight given to IC-2 by Firm #2. Average subjective weights for all firms did show that IC-3 was considered the most important followed by

TABLE 4-14

VALUE OF SUBJECTIVE WEIGHTS BY AUDITOR

Subj - Firm	Subjective Weight for IC Variable			Subj - Firm	Subjective Weight for IC Variable		
	IC-1	IC-2	IC-3		IC-1	IC-2	IC-3
1 - 1	20	10	70	31 - 3	50	25	25
2 - 1	15	35	50	32 - 3	20	10	70
3 - 1	5	70	25	33 - 3	20	40	40
4 - 1	5	45	50	34 - 3	5	25	70
5 - 1	35	45	20	35 - 4	20	35	45
6 - 1	60	0	40	36 - 4	30	20	50
7 - 1	10	30	60	37 - 4	25	50	25
8 - 1	30	35	35	38 - 4	35	40	25
9 - 1	20	30	50	39 - 4	50	20	30
10 - 1	10	15	75	40 - 4	20	35	45
11 - 1	0	0	100	41 - 4	10	20	70
12 - 1	20	30	50	42 - 4	15	30	55
13 - 2	15	40	45	43 - 4	35	25	40
14 - 2	15	35	50	44 - 4	35	15	50
15 - 2	30	20	50	45 - 4	20	30	50
16 - 2	20	35	45	46 - 4	15	15	70
17 - 2	25	30	45	47 - 4	10	10	80
18 - 2	20	70	10	48 - 4	20	40	40
19 - 2	20	30	50	49 - 4	20	30	50
20 - 2	35	30	35	50 - 4	40	20	40
21 - 2	30	20	50	51 - 4	40	20	40
22 - 2	10	25	65	52 - 4	30	20	50
23 - 3	20	35	45	53 - 4	30	30	40
24 - 3	10	40	50	54 - 4	25	25	50
25 - 3	30	20	50	55 - 4	60	20	20
26 - 3	45	35	20	56 - 4	20	30	50
27 - 3	20	40	40	57 - 4	20	20	60
28 - 3	15	35	50	58 - 4	20	30	50
29 - 3	40	20	40	59 - 4	10	50	40
30 - 3	25	50	25	60 - 4	10	40	50

TABLE 4-14 - continued

Subj - Firm	Subjective Weight for IC Variable			Subj - Firm	Subjective Weight for IC Variable		
	IC-1	IC-2	IC-3		IC-1	IC-2	IC-3
61 - 4	30	20	50	85 - 4	60	15	25
62 - 4	10	50	40	86 - 4	5	0	95
63 - 4	5	5	90	87 - 4	10	30	60
64 - 4	5	40	55	88 - 4	75	20	5
65 - 4	25	50	25	89 - 4	30	20	50
66 - 4	20	40	40	90 - 4	20	40	40
67 - 4	35	40	25	91 - 4	40	10	50
68 - 4	5	65	30	92 - 4	30	20	50
69 - 4	10	40	50	93 - 4	30	10	60
70 - 4	10	20	70	94 - 4	20	30	50
71 - 4[a]				95 - 4	10	15	75
72 - 4	30	20	50	96 - 4	10	20	70
73 - 4	15	70	15	97 - 4	25	35	40
74 - 4	10	30	60	98 - 4	10	20	70
75 - 4	0	35	65	99 - 4	30	40	30
76 - 4	33	33	34	100 - 4[a]			
77 - 4	20	20	60	101 - 4	50	15	35
78 - 4	0	10	90	102 - 4[a]			
79 - 4	50	20	30	103 - 4	40	40	20
80 - 4	10	35	55	104 - 4	20	40	40
81 - 4	20	50	30	105 - 4	0	0	100
82 - 4	10	40	50	106 - 4	30	20	50
83 - 4	20	40	40	107 - 4	0	5	95
84 - 4	20	35	45	108 - 4	35	15	50
				109 - 4	50	25	25
				Overall	23.0	28.9	48.1

[a]These subjects did not respond to this question in the post-experiment questionaire.

TABLE 4-15

VALUE OF SUBJECTIVE WEIGHTS BY FIRM AND EXPERIENCE LEVEL

Firm or Experience Level	Average Subjective Weights for Internal Controls		
	IC-1	IC-2	IC-3
Firm			
Firm # 1	19.2	28.8	52.1
Firm # 2	22.0	33.5	44.5
Firm # 3	25.0	31.3	43.8
Firm # 4	23.4	28.0	48.6
Exp.			
E-1 (1-2 years)	24.2	27.6	48.2
E-2 (3 or more)	21.9	30.2	47.8
X-1 (1-3 years)	23.7	28.7	47.6
X-2 (4 or more)	20.5	29.8	49.5
Overall	23.0	28.9	48.1

IC-2 and IC-1, in that order. These results are similar to those for the statistical weights that are presented in Table 4-8.

The average subjective weights by experience level indicate general agreement as to the relative importance of IC-3 over the other controls. Although IC-2 and IC-1 are considered second and third in order of importance for all experience levels, more experienced auditors seemed to attach greater importance to IC-2 and less importance to IC-1 than did the lesser experienced auditors. There is no ready explanation as to why auditors with more experience would either attach greater significance to shipment authorization (IC-2) and/or less significance to credit approval (IC-1) for this audit situation.

Self-insight index. In order to more thoroughly examine the relationship between the subjective weights discussed in the previous section and the statistical weights determined from the ANOVA analysis, a self-insight index was computed for each auditor. Since each auditor made both reliability and sample size judgments, two self-insight indices were computed for each auditor for which ANOVA was possible.

The self-insight indices resulted from calculating the Pearson product-moment correlation measures between the subjective and statistical weights across the three key IACs. As discussed previously, the statistical weights used were "adjusted" weights that included interaction effects.[10] Caution is recommended in evaluating these

[10] See p. 55 above for an explanation of the method used to arrive at the "adjusted" statistical weights.

results since these correlations were computed over only three variables and may therefore be considered unstable.

Table 4-16 shows the self-insight indices for each auditor. These indices vary considerably from a low of -1.0 to a high of +1.0. The self-insight index was much higher for the reliability judgment than for the sample size judgment. For example, 51.5% of the auditors had a reliability index greater than .90. This percentage increased to 63.6% for those with an index greater than .80. For the sample size judgment, only 38.8% of the auditors had an index greater than .90, with the percentage increasing to 52.0% for those with an index greater than .80.

The overall mean self-insight indices were .68 and .59 for the reliability and sample size judgments, respectively. These results are not as high as most studies involving auditor judgment. For example, Ashton [1974] reported a mean Pearson correlation of .89, and extensions of his study obtained mean correlations of .77 [Ashton and Kramer, 1980] and .86 [Ashton and Brown, 1980]. Hamilton and Wright [1977] and Messier [1979] showed results which also indicated a high degree of self-insight. The results in the present study are most similar to those of Joyce [1976] who reported a mean self-insight of .53 with a range of -.78 to 1.00.

The self-insight indices also were summarized by firm and experience levels as shown in Table 4-17. The participants from Firm #1 showed substantially greater self-insight than those from other firms. A possible explanation lies in the fact that nine of the eleven auditors from Firm #1 for which self-insight indices were

TABLE 4-16

SELF-INSIGHT INDICES FOR INDIVIDUAL AUDITORS

Auditor - Firm	Self-Insight Index		Auditor - Firm	Self-Insight Index	
	Reliability	Sample Size		Reliability	Sample Size
1 - 1	a	.99	31 - 3	a	.99
2 - 1	.88	.85	32 - 3	.99	.99
3 - 1	.63	.56	33 - 3	.98	.75
4 - 1	.98	.62	34 - 3	.98	.91
5 - 1	.98	.99	35 - 4	.12	-.31
6 - 1	.90	.74	36 - 4	.97	.99
7 - 1	.95	.99	37 - 4	.45	.91
8 - 1	.77	.47	38 - 4	.76	.76
9 - 1	.98	.96	39 - 4	.76	.76
10 - 1	a	a	40 - 4	.89	.82
11 - 1	1.00	a	41 - 4	.99	.99
12 - 1	.97	.96	42 - 4	.98	.97
13 - 2	.95	.52	43 - 4	.69	-.32
14 - 2	.86	.99	44 - 4	.94	.52
15 - 2	-.76	-.76	45 - 4	.98	.31
16 - 2	.74	.99	46 - 4	.99	.99
17 - 2	.97	.99	47 - 4	.99	a
18 - 2	.99	.99	48 - 4	.87	.62
19 - 2	.96	.98	49 - 4	.95	.96
20 - 2	.27	.18	50 - 4	-.19	.02
21 - 2	.91	.76	51 - 4	.39	.67
22 - 2	.82	.86	52 - 4	.98	.96
23 - 3	.79	.81	53 - 4	.95	.84
24 - 3	.94	.99	54 - 4	.76	-.50
25 - 3	-.35	.27	55 - 4	.65	-.68
26 - 3	.97	.11	56 - 4	.99	.92
27 - 3	.99	.85	57 - 4	-.12	.98
28 - 3	.95	.98	58 - 4	-.97	.53
29 - 3	.71	-.05	59 - 4	.99	.50
30 - 3	.07	-.22	60 - 4	.87	.86

[a]ANOVA results did not allow for determination of statistical weights.

TABLE 4-16 - continued

Auditor - Firm	Self-Insight Index		Auditor - Firm	Self-Insight Index	
	Reliability	Sample Size		Reliability	Sample Size
61 - 4	-.31	-.70	86 - 4	a	a
62 - 4	.85	.79	87 - 4	.94	.93
63 - 4	a	a	88 - 4	.31	.52
64 - 4	.78	.77	89 - 4	.85	.38
65 - 4	-.31	-.100	90 - 4	.96	.70
66 - 4	.65	.68	91 - 4	.80	.93
67 - 4	.79	.77	92 - 4	.94	-.14
68 - 4	.99	.99	93 - 4	.87	.27
69 - 4	.95	.87	94 - 4	.93	.97
70 - 4	a	a	95 - 4	.99	.99
71 - 4	b	b	96 - 4	.40	.99
72 - 4	.96	.94	97 - 4	-.68	.87
73 - 4	1.00	-1.00	98 - 4	-.10	-.92
74 - 4	.99	.97	99 - 4	-.52	.89
75 - 4	.86	.89	100 - 4	b	b
76 - 4	.50	.50	101 - 4	.99	.71
77 - 4	.99	.71	102 - 4	b	b
78 - 4	.99	a	103 - 4	.75	.64
79 - 4	.06	.14	104 - 4	.47	.60
80 - 4	.94	.99	105 - 4	a	a
81 - 4	.99	.99	106 - 4	.94	.95
82 - 4	-.98	-.99	107 - 4	.99	.86
83 - 4	.75	.99	108 - 4	.99	.98
84 - 4	.99	.81	109 - 4	-.50	-.16
85 - 4	.99	.94			
			Overall	.68	.59

[a]ANOVA results did not allow for determination of statistical weights.

[b]Auditor did not provide the necessary subjective weights.

TABLE 4-17

SELF-INSIGHT INDICES BY FIRM AND EXPERIENCE LEVEL

Firm or Experience Level	Average Value of Self-Insight Index	
	Reliability	Sample Size
Firm # 1	.90	.81
Firm # 2	.67	.65
Firm # 3	.73	.62
Firm # 4	.64	.55
Exp.		
E-1 (1-2 years)	.62	.55
E-2 (3 or more)	.74	.64
X-1 (1-3 years)	.65	.56
X-2 (4 or more)	.78	.73
Overall	.68	.59

calculated had 4 or more years of experience and all were participants in an advanced statistical sampling seminar at the time of the current study. When all firms are considered and either experience classification scheme is used, the more experienced auditors had higher self-insight indices for both the reliability and sample size judgments. For the reliability judgment, the more experienced auditors (E-2 and X-2) had indices of .74 and .78 as compared to .62 and .65 for the lesser experienced auditors (E-1 and X-1). Similar results are shown for the sample size judgment where indices of .64 and .73 were obtained for the more experienced auditors as compared to .55 and .56 for the lesser experienced auditors. Results from the studies conducted by Ashton [1974], Hamilton and Wright [1977], Ashton and Brown [1980], and Ashton and Kramer [1980] indicate that judgment insight increases with increased levels of auditing experience. Greater familiarity with the task provides a possible explanation for the superior insight of those auditors with at least 3 years of experience.

Additional Data

Additional data is available to better understand the judgments of the auditors participating in the current study. First, results are provided to indicate the approach taken by the auditors in using the statistical sampling tables in arriving at the sample size decision. Summary comments are then presented to reflect the responses to questions in the post-experiment questionnaire.

Decision making approach. The selection of a sample size from the statistical tables required the auditors to choose an

appropriate materiality and confidence level for each case. The materiality and confidence levels chosen by each auditor across the 12 cases were examined and are summarized in Table 4-18. Again, the familiarity of statistical sampling techniques by the participants in Firm #1 may underlie their decision to maintain a constant materiality across all cases. This approach can be considered the most appropriate in the application of a statistical sampling plan.

The approach taken by most of the auditors from Firm #3 and Firm #4 can be contrasted to that approach taken by members of Firm #1. Most of these auditors chose to vary both the materiality and confidence level across the 12 cases. No strong firm preference was revealed by members of Firm #2. A possible explanation for these results can be found by examining some results of the answers to the post-experiment questionnaire.

Questionnaire results. The post-experiment questionnaire provided the opportunity for the auditors to respond in some important areas. Three specific areas of interest were the representiveness of the task, the auditors' familiarity with the task, and the extent to which similar decisions are made by the auditor. Overall, the task was considered "moderately representative," receiving an average rating of 3.0 on a 5-point scale. The auditors' familiarity with the task and the extent to which they make similar decisions also were rated by the auditors. With a rating of 3 representing "moderately familiar" and a 4 representing "not very familiar," the average rating of task familiarity was 3.25. A contrast between Firm #1 (2.50) and Firm #3 (3.67) again is very much in evidence. Finally, with a 3 representing "sometimes" and a 4 representing "often" with

TABLE 4-18

SAMPLE SIZE DECISIONS - BY FIRM

Decision Basis	Firm #1	Firm #2	Firm #3	Firm #4	Total
Constant Materiality (M)	8	3	1	18	30
Constant Confidence Level (CL)	1	3	2	11	17
Variation in M and CL	3	4	9	44	60
Total	12	10	12	75	109

[a]Two subjects responded with decisions of either 0 or a number that reflects both a constant materiality and confidence level.

respect to the extent of similar decisions, the overall average was 3.04. The experience level and use of statistical sampling by the auditors in Firm #1 is evidenced by their average rating of similar decisions of 3.67, as compared to that of Firm #3 of 3.17. These results provide some explanation as to the differences between responses of Firm #1 and Firm #3 members that were referred to in the previous section.

Eight post-experiment questions were included to reflect the extent of motivation of the auditors in performing the experimental task. The analysis of responses to these questions reflects a substantial difference between those auditors in the lowest quartile (low motivation) and those in the highest quartile (high motivation). A substantial difference is also found between the mean self-insight indices in the highest quartile (.87) and those in the lowest quartile (.48).

Summary

The results of the study were presented and briefly discussed in this chapter. An examination of the responses from the preliminary audit program stage revealed some differences among both the individual auditors and the participating firms. Also, the evaluation of the auditors' responses to the compliance test decisions suggested that IC-3 was considered the most important IAC, with IC-2 and IC-1 following in that order.

Responses provided at the audit program revision stage provided the opportunity for the primary and more extensive analysis. Specifically, ANOVA results were used to evaluate the relative importance of

the key internal controls provided in this specific reliability/sample size judgment. In addition, the extent of consensus among the auditors was examined with results summarized by firm and experience level. Self-insight indices also were calculated and summarized by firm and experience level. Finally, other information pertaining to the decision processes used by the auditors was presented, followed by some results of a post-experiment questionnaire.

Results of analyzing the responses at the preliminary audit program stage are summarized as follows:

1. Differences among the firms were noted with respect to the reliability and sample size responses.
2. Responses to the compliance test decisions suggested that IC-3 was considered the most important control, followed by IC-2 and IC-1, in that order.

Analysis of responses to the twelve case situations presented at the audit program revision stage produced results as follows:

1. IC-3 (control over physical shipment) accounted for most of the variance in the auditors' judgments, with IC-2 (shipment authorization) and IC-1 (credit approval) following in order of importance.
2. Individual auditor responses indicated that approximately 62% of the auditors had most of the variance accounted for by IC-3, approximately 24% of the auditors had IC-2 account for most of the variance in their responses, and the remaining 14% of the auditors had IC-1 account for most of the variance.

3. Significant interaction effects were found for fifteen of the 102 auditors on the reliability and/or sample size decisions.

4. Evaluation of judgment consensus resulted in a mean canonical correlation of .78 with a standard deviation of responses of .145. These results represent reasonable agreement among the auditors, but nevertheless reveal some variability among their responses.

5. Another measure of consensus was obtained from the use of Pearson product-moment correlation. The mean correlation for the reliability judgment was .693 with a standard deviation of .19. The sample size judgment had a mean correlation of .609 with a standard deviation of .22.

6. The degree of self-insight exhibited by the auditors can be considered somewhat low when compared to other accounting studies. The mean values for the reliability and sample size judgments were .68 and .59, respectively.

7. Firm and experience levels were considered in examining the above results. In most cases, firm and/or experience levels were shown to impact on the results.

The following chapter will discuss these results in greater detail in conjunction with some implications, suggestions for future research, and conclusions.

CHAPTER V

SUMMARY, CONCLUSIONS, AND SUGGESTED FUTURE RESEARCH

Summary of the Problem and Research Approach

This dissertation has provided additional evidence to facilitate a better understanding of the role of auditor judgment in the evaluation of IAC information and the decision as to the extent of subsequent audit procedures (substantive tests). A laboratory experiment was conducted to gather data which were used to examine auditor judgment about the reliability of an IAC system and the resulting selection of a sample size for a substantive test. The participants were 109 members of four "Big-Eight" accounting firms. A descriptive approach was taken within the framework provided by the Brunswik lens model to examine these decisions for a specific audit task.

Although other studies have examined similar decisions made by auditors, some specific and significant limitations were thought to exist in these studies. The present study sought to overcome those limitations. Primarily, a more realistic setting was used, with a greater degree of control over the significant factors impacting upon the decisions of the auditors. These improvements were discussed in detail in Chapter II of this dissertation.

Two key decision points within the audit process that require the exercise of professional judgment were examined in this study. These included: (1) the preliminary audit planning stage and (2) the

audit program revision stage. Examination of the responses from these stages resulted in descriptions of the auditors' judgments and the evaluation of auditor consensus and self-insight. However, the extent of auditor consensus is the focal point in this study for examining the exercise of professional judgment within the audit process.

There are a number of reasons to expect and/or hope for a certain level of consensus among auditors who are applying their professional judgment to the same audit task. First, most auditors have similar educational backgrounds in accounting. In addition, accounting firms provide additional training that lends itself to a common basis for decision making among the auditors. Finally, to become a Certified Public Accountant an individual is required to pass a uniform exam by providing responses that indicate adherence to a prescribed set of standards established by the accounting profession. Therefore, it appears that the lack of consensus among auditors is considered more costly than the efforts to achieve an accepted level of agreement.

Lack of consensus among auditors can be considered costly for a number of reasons. A lack of consensus could suggest that the decisions of some auditors are either resulting in unnecessary and costly audit work or that the firm is subjecting itself to additional risk (cost) as a result of insufficient audit work. Additionally, the auditors who are making less than "optimal" decisions are sharing their "expertise" with others in the firm since a large degree of learning in auditing is done "on the job." Finally, lack of consensus could lead to excessive and costly review procedures that would be required in addition to those that are normally present. The description of the auditors' judgments and the computation of self-insight

indices are related directly to the topic of consensus. Results from this study in all of these areas are presented in the following section.

Summary of the Results of the Study

A word of caution should be presented prior to discussing the results of the current study. Weber [1978, p. 384] notes that any findings derived from a fixed effects model can not be generalized. For this reason, and others that were enumerated earlier, generalization of these results beyond the participating auditors and the specific task should be avoided. However, since a realistic setting was provided as representative of the type of reliability and sample size decisions that are commonly made by auditors, meaningful statements can still be made from evaluating the results of the study.

Description of Auditors' Judgments

The description of the auditors' decisions both at the preliminary audit program stage and at the audit program revision stage indicated that the most important control in deciding that recorded sales were for valid transactions was one concerning control over physical shipment (IC-3), followed by the controls for shipment authorization (IC-2) and credit approval (IC-1), respectively. The fact that the least important control (IC-1) accounted for 24.9% of the variance in significant ($\alpha \leq .05$) sample size judgments and 11.4% of the variance in all sample size judgments made at the program revision stage is an indication that each of the internal controls was considered important. IC-3 was recognized as the most important control irrespective of the firm or experience level chosen for the reliability and sample size

decisions, except for the sample size decisions of Firm #2 where IC-2 was considered slightly more important. The importance of these controls, as discussed above, must be considered the "relative" importance within the context of the current study.

The Extent of Judgment Consensus

The extent of consensus among auditors at the audit program revision stage was evaluated through the use of canonical correlation, Pearson product-moment correlation, and cluster analysis. The mean canonical correlation was high with a value of .78, but also showed high variability with a standard deviation of .145. The mean Pearson product-moment correlations for the reliability and sample size judgments were .69 and .61, respectively. The standard deviations for these judgments were again quite high at .19 and .22, respectively. These results indicate a limited amount of consensus among the participating auditors. Analysis of firm and experience levels showed these factors to have a significant impact on the extent of consensus among auditors. The cluster analysis results were inconclusive as to the extent of consensus among the auditors or as to the impact of firm or experience differences.

In commenting on the level of consensus found in this study there are two very important issues that should be addressed. First, the reasons for any lack of consensus must be examined. Second, the costs associated with the various levels of consensus must be known and properly evaluated.

While the level of consensus obtained in this study appears low, the relative complexity of the judgment task must be considered a

contributing factor. The absence in this study of any review process or joint decision making that would be found in actual practice is another possible explanation. Other possible reasons can be found in the results that indicated there were differences among firms and between experience levels. That is, different auditing philosophies or the training provided by the individual firms was evident. Also, the extent of an individual's practical experience impacted on the extent of consensus.

The current descriptive study was not concerned with examining the costs of various levels of consensus as a basis for establishing this as a major problem within the auditing profession. Any implications from the results presented above with respect to the extent of consensus among the participating auditors are difficult to discern. As discussed in the previous section, there are many indications that the auditing profession desires a certain level of consensus among its members. However, the results of this study present no indication of a need for "corrective action" from individual accounting firms or the accounting profession to increase the current level of consensus.

The Degree of Self-Insight

A comparison of the auditor's subjective evaluation of the relative importance of the three key IACs and the importance as reflected in the statistical weights was obtained through a self-insight index. The overall mean self-insight indices were .68 and .59 for the reliability and sample size judgments, respectively. Greater self-insight was reflected by those with more years of

experience. The ability to understand and express a decision approach is most important for the auditors participating in this study. As seniors with in-charge responsibilities, the auditors have responsibility for communication skills both to their subordinates and to their superiors. An extensive amount of learning takes place "on the job" and provides the auditor with a basis for formulating an audit decision approach. For these reasons, the low self-insight indices in the current study are a source of some concern. The fact that the auditors were responding to a relatively complex judgment task is a possible explanation for the somewhat low indices calculated in this study.

Implications for the Auditing Profession

In comparing this study with similar auditing studies,[1] results substantiate previous findings in some areas and provide some new and meaningful findings in other areas. As an important extension of earlier studies, this study has examined a greater portion of the audit revision process and in greater detail than has been done before. A significant finding in this study was some evidence that differences existing among auditors and/or firms at the preliminary audit program stage appear to continue to exist at the audit program revision stage. Of course, this finding requires additional research support prior to suggesting any immediate implications.

---

[1]These studies would include Ashton [1974], Joyce [1976], Mock and Turner [1979] and others referred to in Chapter II above.

Although previous studies have used various audit tasks, the level of consensus found in this study was comparable to that obtained in other audit studies. The use of statistical sampling techniques and a rather complex audit task enhanced the contribution of this study. The auditing profession should be most interested in the impact of introducing additional realism into these types of studies and evaluating the results. In addition, information as to the "perceived" importance of key internal controls in examining various other audit objectives can be obtained through the study of similar audit tasks.

The various approaches taken by auditors to determine a statistical sample can provide meaningful information to auditors, firms, and to the auditing profession. Findings in this study indicate that differences do exist among auditors in their application of statistical sampling techniques. To the extent that statistical sampling is a significant portion of the decision making aspect of the audit process, the ability of auditors to use this tool should be appropriately monitored.

In summary, perhaps the primary implication from this study is that it is not obvious that the level of consensus desired by the auditing profession is being achieved. Of course, there were indications that more consensus was being achieved within the firms and at more experienced levels, which may be considered more important than across the entire auditing profession. A major concern is whether or not resources are being properly utilized to enhance the probability of consensus among auditors. For this reason, the auditing profession must be aware of any evidence that suggests the current level of

consensus among auditors is not optimal with respect to the utilization of available resources.

Additionally, perhaps both the extent of consensus and self-insight could be enhanced if more attention was given to providing feedback to the auditors with respect to their "on the job" decisions. As mentioned in the previous section a first step would be to determine the reasons for such differences, and then to perform a cost-benefit analysis to allow for the comparison of alternatives. With "on the job" feedback both difficult and costly, additional consideration should be given to the approach currently taken in firm training sessions to provide meaningful feedback.

The current study is one of many recent studies that must be built upon and extended to provide the necessary impetus for systematically examining the audit process. In recent years members of the auditing profession and academic researchers have come to realize that the complexity of the audit process does not preclude it from being subjected to appropriate research examination. Such suggestions for future research relating to the current study are provided in the following section.

Suggestions for Future Research

Since this study was not intended to be an end in itself, a number of possible avenues for future research exist in this area. Many opportunities exist to examine other portions of the audit process that would involve use of different internal accounting controls and/or audit objectives. Other adjustments in a similar study would

include the manner in which information is provided to the auditors or the types of decisions required. Elicitation of subjective probabilities seems to be a viable alternative approach to examine the decision making process of auditors. Also, the fact that auditors are involved in the sequential processing of information contributes to the attractiveness of a Bayesian approach.

Audit decisions are also made at various levels within the accounting firm. Use of managers or partners in similar audit situations might provide greater insight into the extent of consensus on "final" decisions versus those made "up front" by those auditors with considerably less experience. Similar research could be conducted using audit teams as subjects which incorporates a more realistic group decision making situation and would incorporate the possibility of some type of peer review. Incorporation of such a realistic environment, however, can lead to many additional research problems not faced when using individual responses.

A broader approach that could be used to examine this problem area would include an attempt to determine what factors seem to enhance consensus among auditors. This information, used in conjunction with information about costs associated with attaining consensus versus the costs for unacceptable low levels of consensus, could affect some standard operating procedures within the auditing profession.

A final area that warrants additional auditor judgment research deals with the development of an operational model that might better establish the relationship between internal accounting control

evaluation and the resulting extent of substantive testing. The testing of such a model to determine its usefulness in audit decision making situations would also be warranted. This type of research would lend itself to what has been reflected as a major concern and premise of this study, that is, inefficient allocation of resources within an audit.

The role of professional judgment has been and will continue to be a major part of the audit process. However, the need to continue to examine this process and to expose it to systematic and meaningful research will not diminish in the near future. This research has been presented with the hope that it can in some way serve as a stepping stone for future research in this vitally important area.

BIBLIOGRAPHY

American Accounting Association. "Report of the 1976-77 Committee on Human Information Processing." AAA, August 1977.

American Institute of Certified Public Accountants. Commission on Auditors' Responsibilities: Report, Conclusions, and Recommendations. New York: AICPA, 1978.

American Institute of Certified Public Accountants. Professional Standards - Auditing, Management Advisory Services, Tax Practice, Vol. 1. Chicago: Commerce Clearing House, 1979a.

American Institute of Certified Public Accountants. Report of the Special Advisory Committee on Internal Accounting Control. New York: AICPA, 1979b.

Arens, A. A., and J. K. Loebbecke. Auditing: An Integrated Approach. Englewood Cliffs: Prentice-Hall, Inc., 1976.

Arkin, H. "Statistical Sampling and Internal Control." The CPA Journal (January 1976), pp.15-18.

Ashton, R. H. "Judgment Formation in the Evaluation of Internal Control: An Application of Brunswik's Lens Model." Ph.D. dissertation, University of Minnesota, 1973.

Ashton, R. H. "An Experimental Study of Internal Control Judgments." Journal of Accounting Research (Spring 1974), pp. 143-57.

Ashton, R. H. "The Predictive-Ability Criterion and User Prediction Models: A Reply." The Accounting Review (July 1976), pp. 680-82.

Ashton, R. H. "Some Implications of Parameter Sensitivity Research for Judgment Modeling in Accounting." The Accounting Review (January 1979a), pp. 170-79.

Ashton, R. H. "Comment: Some Observations on Auditors' Evaluations of Internal Accounting Controls." Journal of Accounting, Auditing and Finance (Fall 1979b), pp. 56-66

Ashton, R. H., and P. R. Brown. "Descriptive Modeling of Auditors' Internal Control Judgments." Forthcoming in Journal of Accounting Research (1980).

Ashton, R. H., and S. L. Kramer. "Students As Surrogates in Behavioral Accounting Research: Some Evidence." Forthcoming in Journal of Accounting Research (1980).

Bernstein, L. A. "The Concept of Materiality." The Accounting Review (January 1967), pp. 86-95.

Bodnar, G. "Reliability Modeling of Internal Control Systems." The Accounting Review (October 1975), pp. 747-57.

Boer, G. "The Role of Judgment in Statistical Sampling." The CPA Journal (March 1974), pp. 39-43.

Brehmer B. "Note on Clinical Judgment and the Formal Characteristics of Clinical Tasks." Psychological Bulletin (September 1976), pp. 778-82.

Brunswik, E. "Organismic Achievement and Environmental Probability." Psychological Review (May 1943), pp. 255-72.

Brunswik, E. "Systematic and Representative Designs of Psychological Experiments." In V. Neyman (Ed.), Berkeley Symposium on Mathematical Statistics and Probabilities. Berkeley: University of California Press, 1949.

Brunswik. E. Conceptual Framework of Psychology. Chicago: University of Chicago Press, 1952.

Burns, D. "Extending the Study and Evaluation of Internal Control." The CPA Journal (May 1974), pp. 31-35.

Burns, D., and J. K. Loebbecke. "Internal Control Evaluation: How the Computer Can Help." The Journal of Accountancy (August 1975), pp. 60-70.

Casey, C. J., Jr. "The Predictive-Ability Criterion and User Prediction Models: A Comment." The Accounting Review (July 1976), pp. 677-79.

Castellan, N. J., Jr. "The Analysis of Multiple Criteria in Multiple-Cue Judgment Tasks." Organizational Behavior and Human Performance (October 1972), pp. 242-61.

Cook, J. M., and T. P. Kelley. "Internal Accounting Control: A Matter of Law." The Journal of Accountancy (January 1979), pp. 56-64.

Corless, J. C. "Assessing Prior Distributions for Applying Bayesian Statistics in Auditing." The Accounting Review (July 1972), pp. 556-66.

Cushing, B. "A Mathematical Approach to the Analysis and Design of Internal Control Systems." The Accounting Review (January 1974), pp. 24-41.

Dawes, R. M. "The Mind, the Model, and the Task." In F. Restle, R. M. Shiffring, N. J. Castellan, H. R. Lindman and D. B. Pisoni (Eds.), Cognitive Theory, Vol. 1. Hillsdale: Lawrence Erlbaum Associates, 1975, pp. 119-29.

Dawes, R. M. and B. Corrigan. "Linear Models in Decision Making." Psychological Bulletin (February 1974), pp.95-106.

Deming, W. E. "On a Problem in Standards of Auditing From the Viewpoint of Statistical Practice." Journal of Accounting, Auditing and Finance (Spring 1979), pp. 197-208.

Edwards, W. "Bayesian and Regression Models of Human Information Processing - A Myopic Perspective." Organizational Behavior and Human Performance (November 1971), pp. 639-48.

Einhorn, H. J. "Expert Judgment: Some Necessary Conditions and an Example." Journal of Applied Psychology (October 1974), pp. 562-71.

Einhorn, H. J. "Overconfidence in Judgment." In R. A. Shweder and D. W. Fiske (Eds.), New Directions for Methodology of Behavioral Research: Fallible Judgment in Behavioral Research. San Francisco: Jossey-Bass, 1980.

Einhorn, H. J., and R. M. Hogarth. "Confidence in Judgment: Persistence of the Illusion of Validity." Psychological Review, Vol. 85, No. 5 (1978), pp.395-416.

Elliott, R. K., and J. R. Rogers. "Relating Statistical Sampling to Audit Objectives." The Journal of Accountancy (July 1972), pp. 46-55.

Felix, W. L., and J. L. Goodfellow. "Audit Tests for Internal Control Reliance," Working paper, University of Washington, 1978.

Foreign Corrupt Practices Act of 1977, amend. Sec. 13(b) of the Securities Exchange Act of 1934 (15 U.S.C. 78q(b)).

Gibbins, M. "Human Inference, Heuristics and Auditors' Judgment Processes." Presented at the C.I.C.A. Auditing Research Symposium, Laval University, November 1977.

Goldberg, L. R. "Simple Models or Simple Processes?" American Psychologist (July 1968), pp. 483-96.

Grimlund, R. A. "The Integration of Internal Control System and Account Balance Evidence." Working paper, University of Iowa, August 1978.

Hamilton, R. E., and W. F. Wright. "The Evaluation of Internal Controls Over Payroll." Graduate School of Business Research Paper No. 397. Stanford University, 1977.

Hofstedt, T. R., and C. D. Hughes. "An Experimental Study of the Judgment Element in Disclosure Decisions." The Accounting Review (April 1977), pp. 379-95.

Johnson, S. C. "Hierarchical Clustering Schemes." Psychometrika, XXIII (1967), pp. 241-54.

Joyce, E. J. "Expert Judgment in Audit Program Planning." Studies on Human Information Processing in Accounting: 1976. Supplement to Journal of Accounting Research, 14, pp. 29-60.

Kaplan, R. S. "The Roles for Research and Development in Auditing." Presented at the Symposium on Auditing Research. University of Illinois, 1977, pp. 3-11

Kennedy, H. A. "Human Information Processing in Auditing: The Lens Model and Auditors' Judgments." Prepared for the C.I.C.A. Auditing Research Symposium. Laval University, November 1977.

Kinney, W. "A Decision Theory Approach to the Sampling Problem in Auditing." Journal of Accounting Research (Spring 1975a), pp. 117-32.

Kinney, W. "Decision Theory Aspects of Internal Control System Design/ Compliance and Substantive Tests." Studies on Statistical Methodology in Auditing: 1975b. Supplement to Journal of Accounting Research, 13, pp. 14-37.

Konrath, L. F. "The CPA's Risk in Evaluating Internal Control." The Journal of Accountancy (October 1971), pp. 53-56.

Libby, R. "Man Versus Model of Man: The Need for a Nonlinear Model." Organizational Behavior and Human Performance (June 1976a), pp. 23-26.

Libby, R. "Man Versus Model of Man: Some Conflicting Evidence." Organizational Behavior and Human Performance (June 1976b), pp. 1-12.

Libby, R., and B. L. Lewis. "Human Information Processing Research in Accounting: The State of the Art." Accounting, Organizations and Society, Vol. 2, No. 3, 1977, pp. 245-68.

Lin, W. T., J. J. Mock, L. K. Newton, and M. A. Vasarhelyi. "A Review of Audit Research." Working paper, University of Southern California, March 1978.

Loebbecke, J. K., and J. Neter. Considerations in Choosing Statistical Methodology in Auditing: 1975. Supplement to Journal of Accounting Research, 13, pp. 38-69.

Mautz, R. K., "Evidence, Judgment and the Auditor's Opinion." The Journal of Accountancy (April 1959), pp. 40-44.

Mautz, R. K., and D. L. Mini. "Internal Control Evaluation and Audit Program Modification." The Accounting Review (April 1966), pp. 283-91.

Mautz, R. K., and H. A. Sharaf. The Philosophy of Auditing. Sarasota: American Accounting Association, 1961.

Messier, W. F., Jr. "An Examination of Expert Judgment in the Materiality/Disclosure Decision." Ph.D. dissertation, Indiana University, 1979.

Mock, T. J., and J. L. Turner. "The Effect of Changes in Internal Controls on Audit Programs." In T. J. Burns (Ed.), Behavioral Experiments in Accounting II, 1979.

Morris, W., and H. Anderson. "Audit Scope Adjustments for Internal Controls?" The CPA Journal (July 1976), pp. 15-20.

Nunnally, J. C. Psychometric Theory. New York: McGraw Hill, Inc., 1967.

Peat, Marwick, Mitchell & Co. Research Opportunities in Auditing. New York: Peat, Marwick, Mitchell & Co., 1976.

Reckers, P., and M. Taylor. "Consistency in Auditors' Evaluations of Internal Accounting Controls." Journal of Accounting, Auditing and Finance (Fall 1979), pp. 42-55.

Roberts, D. M. Statistical Auditing. New York: American Institute of Certified Public Accountants, 1978.

Robertson, J. C. Auditing. Dallas: Business Publications, Inc., 1979.

Slovic, P. "Analyzing the Expert Judge: A Descriptive Study of Stockbrokers' Decision Processes." Journal of Applied Psychology, 80 (1969), pp. 255-63.

Slovic, P., B. Fischhoff, and S. Lichtenstein. "Behavioral Decision Theory." In M. R. Rosenzweig and L. W. Porter (Eds.), Annual Review of Psychology. Annual Reviews, 1977, pp. 1-39.

Slovic, P., D. Fleissman, and W. S. Bauman. "Analyzing the Use of Information in Investment Decision Making: A Methodological Proposal." Journal of Business (April 1972), pp. 283-301.

Slovic, P., and S. C. Lichtenstein. "Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment." Organizational Behavior and Human Performance (November 1971), pp. 649-744.

Smith, K. A. "The Relationship of Internal Control Evaluation and Audit Sample Size." The Accounting Review (April 1972), pp. 260-69.

Snedecor, G. W., and W. G. Cochran. Statistical Methods. Ames: Iowa State University Press, 1967.

Swieringa, R., M. Gibbins, L. Larsson, and J. Sweeney. "Experiments in the Heuristics of Human Information Processing." Empirical Research in Accounting: Selected Studies: 1976. Supplement to Journal of Accounting Research, 14, pp. 159-87.

Teitlebaum, A. D., and C. F. Robinson. "The Real Risks in Audit Sampling." Studies on Statistical Methodology in Auditing: 1975. Supplement to Journal of Accounting Research, 13, pp. 70-97.

Turner, J. L., and T. J. Mock. "Economic Considerations in Designing Audit Programs." The Journal of Accountancy (March 1980), pp. 65-74.

Tversky, A., and D. Kahneman. "Judgment Under Uncertainty: Heuristics and Biases." Science, 185 (1974), pp. 1124-31.

Uecker, W. C., and W. R. Kinney, Jr. "Judgmental Evaluation of Sample Results: A Study of the Type and Severity of Errors Made by Practicing CPAs." Accounting, Organizations and Society, Vol. 2, No. 3, 1977, pp. 269-75.

Warren, C. S. "Audit Risk." The Journal of Accountancy (August 1979), pp. 66-74.

Weber, R. "Auditor Decision Making on Overall System Reliability: Accuracy, Consensus, and the Usefulness of a Simulation Decision Aid." Journal of Accounting Research (Autumn 1978), pp. 368-88.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw Hill, Inc., 1971.

Wright, W. F. "Discussion of Expert Judgment in Audit Program Planning." Studies on Human Information Processing in Accounting: 1976. Supplement to Journal of Accounting Research, 14, pp. 61-67.

Wright, W. F. "Self-Insight into the Cognitive Processing of Financial Information." Accounting, Organizations and Society, Vol. 2, No. 4, 1977, pp. 323-31.

Yu, S., and J. Neter. "A Stochastic Model of the Internal Control Systems." Journal of Accounting Research (Autumn 1973), pp. 273-95.

APPENDIX

THE EXPERIMENTAL MATERIALS

SCHOOL OF ACCOUNTING • UNIVERSITY OF FLORIDA • GAINESVILLE 32611
904-392-0155

Dear Participant:

I would like to take this opportunity to thank you for participating in this study. A number of other auditors will also be participating in the study. In reporting the results of this study all individuals and their associated firms will remain anonymous.

Your careful consideration of the information provided and the questions asked will help to ensure reliable and meaningful results.

Again, thank you for your time and assistance in this project.

Sincerely,

Rick Tabor
Ph.D. Candidate

INTRODUCTION

The purpose of this study is to examine auditor judgment in a routine audit situation. Specifically, I am interested in your consideration of the key factors of internal control evaluation and compliance test results in determining the extent of substantive tests to be performed.

Please read carefully the following information and attempt to make your decisions in terms of the setting provided. I realize that the following materials do not include all the information you might like to have for your decisions. However, as with many real situations, it is not always possible to have all the information you would like. For this reason, please respond to the best of your ability on the basis of the information that has been provided.

INSTRUCTIONS

Assume that you have just been assigned as senior-in-charge of the audit of Tabkol Manufacturing Company (TMC). This is the first year that TMC has required an audit and your firm was chosen to provide this audit for the year ended December 31, 1978. Unfortunately for you, the senior-in-charge who had started the preliminary work for this job in September is no longer with your firm. It is now time for the preliminary work to be completed and the year-end work to be planned, and you have been asked to take over at this point.

This study is concerned with planning the amount of compliance and substantive testing that is appropriate within the sales cycle at year-end after evaluating the information available from the preliminary work completed in September. Most importantly, you should be concerned with only one objective while planning your audit program relative to sales. That is, the only audit objective within the context of this study will be that, "recorded sales are for valid transactions."

Most of the information you will be presented is unchanging background information about TMC and other unchanging information gathered during the preliminary audit work. Some of the information about TMC (specifically, three key internal controls whose existence gives assurance concerning the audit objective and compliance test results concerned with these controls) is not held constant. These factors (key internal controls and compliance test results) are varied to form

12 different hypothetical situations you might encounter in your audit of the sales cycle of TMC. You will be asked to provide the appropriate sample sizes for compliance and substantive tests in the preliminary audit program and a final sample size for a substantive test after considering the specific information presented in each case situation along with the unchanging background information.

The Procedures To Follow In This Study Are As Follows:*

(1) Review the information provided to this point (Instructions and Introduction) and seek any clarifications.

(2) Carefully familiarize yourself with the Background Information and Flowchart that follows, and is common to all cases, before proceeding.

(3) Examine the additional information presented for each scenario and answer the questions for the scenario and the appropriate cases before proceeding to the next scenario and doing the same.

*You may review the Introduction, Instructions, Background Information, and/or Flowchart at any time; however, these need not be reviewed for each different scenario or case situation.

BACKGROUND INFORMATION

- Tabkol Manufacturing Company (TMC) is a medium-sized firm which manufacturs hand tools such as hammers and screwdrivers. Its products are distributed nationally through commission-paid representatives who sell building materials, power tools, and other items to hardware stores and building supply outlets. The firm's products are generally displayed on racks by the retailers.

- TMC has experienced an annual growth rate in sales of approximately 10 percent over the past several years. The client currently has approximately 6,000 regular customers. These customers average about eight orders per year (therefore, approximately 48,000 orders are processed yearly), and the average order size is about $200. Thus, sales (all are credit sales) amounted to almost $10 million last year. Accounts receivable comprise nearly 34 percent of the total assets and there is no inventory on consignment.

- The previous auditor anticipated that the balance sheet at December 31, 1978, would closely resemble the following:

Cash	$ 500,000	Accounts payable	$ 700,000
Accounts receivable	1,400,000	Accrued taxes	100,000
Inventory	1,200,000	Bank debt	1,500,000
Plant and equipment	1,000,000	Equity	1,800,000
	$4,100,000		$4,100,000

- Furthermore, the previous auditor anticipated that profit would amount to about $600,000 before taxes on the $10 million in sales.

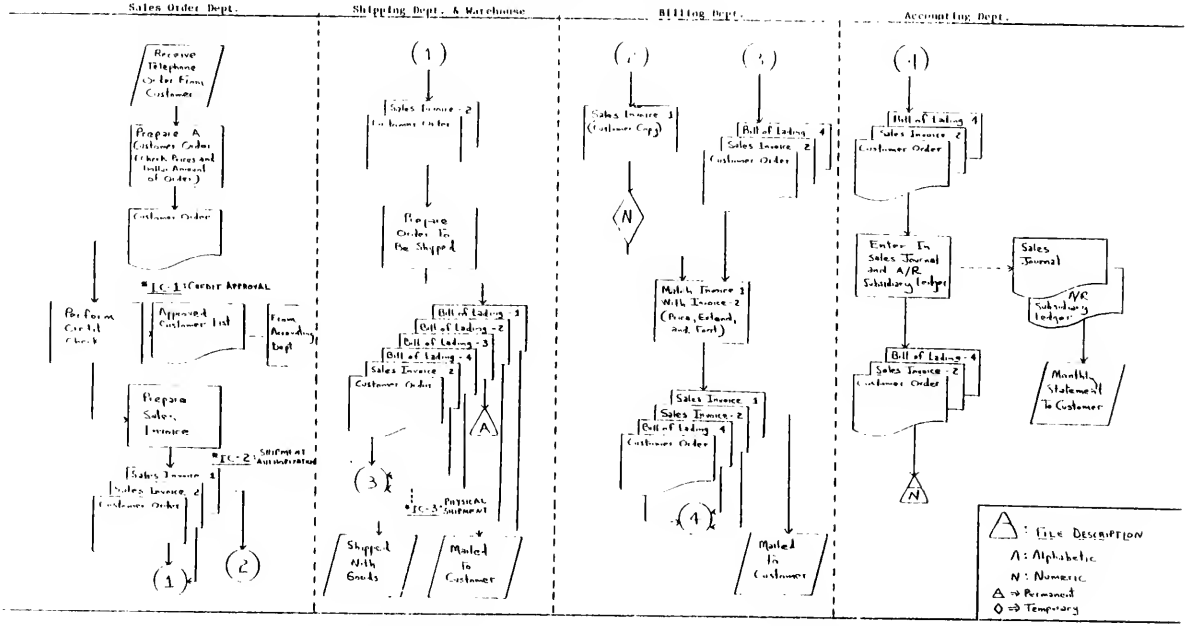
- The accompanying flowchart* reflects the operations of the sales cycle for TMC. Note the representation of the three key internal controls (IC-1, IC-2, and IC-3) within the chart. It is important to remember that the existence of these controls may vary among the different cases.

- Sales returns and allowances for TMC are too immaterial to include in the accompanying flowchart or to verify in the audit.

*The flowchart was constructed as a result of discussions with key personnel.

TARKOL MANUFACTURING COMPANY

- Flowchart Representation of the Sales Cycle



* See the following page for a description of the key internal controls.

Additional Comments on Sales Cycle:

- There are appropriate separation of duties.
- The sales cycle is part of a completely manual system.
- There are no internal audit personnel.

NARRATIVE TO ACCOMPANY FLOWCHART

Telephone orders are received from customers by the sales order department. A clerk prepares a customer order document (checking prices, computing the dollar amount of the order, etc.). Another clerk checks the approved customer list provided by the Accounting Department for credit approval. The customer order is initialed to indicate approval, or if the customer is not on the list, the order is sent directly to the Accounting Department for approval or notice to the customer that prepayment is required.

An invoice is prepared in two copies and the customer order accompanies the shipping copy of the invoice. The customer copy of the invoice (Copy 1) is sent to the billing department and held in the "pending" file awaiting notification that the order was shipped. Notification is indicated by the receipt of the shipping copy of the invoice with the supporting customer order and bill of lading.

The billing clerk matches the received shipping copy with the customer copy from the pending file. Both copies of the invoice are priced, extended, and footed. The customer copy is then mailed directly to the customer and the shipping copy (with the supporting customer order and bill of lading) is sent to the accounts receivable clerk.

Specific Audit Objective: Recorded sales are for valid transactions.

Description of Key Internal Controls: (These may or may not be present in the following case materials.)

IC-1: Procedures exist to control for credit approval as indicated by proper initials on supporting customer order. A customer order is initialed to indicate credit approval if the customer is on the approved credit list provided by the Accounting Department or is approved through the alternative procedures. (See the narrative accompanying the flowchart.)

- Noncompliance implies lack of proper initials on the customer order.

* Expected Noncompliance Rate for IC-1 = 2%

IC-2: Procedures exist to control for shipment authorization as indicated by proper initials on the sales invoice. A sales invoice is initialed to indicate that the sales invoice is in agreement (i.e., as to customer name, quantity of goods, price, etc.) with the customer order and that the shipping department is authorized to make the appropriate shipment.

- Noncompliance implies lack of proper initials on the sales invoice.

* Expected Noncompliance Rate for IC-2 = 2%

IC-3: Procedures exist to control for the physical shipment of goods as indicated by proper initials on the bill of lading. Copy 4 of the bill of lading is initialed to indicate the goods were actually shipped and were supported by a bill of lading (Copy 2) that agreed with the shipment as to the name of the customer and quantity of goods.

- Noncompliance implies lack of proper initials on Copy 4 of the bill of lading.

* Expected Noncompliance Rate for IC-3 = 2%

* These expected noncompliance rates are based upon a review of the Sales Cycle Flowchart and a preliminary assessment of the controls believed to be operating.

NOTE: Other important controls relative to this objective include: 1) the review and follow-up of customer complaints resulting

from the receipt of a monthly statement and 2) the prenumbering of documents. These controls will not be formally tested as it was determined through alternative procedures (i.e., observation of and discussions with key personnel and a general review of the documents) that the few complaints that are received are very infrequent and not material and that all documents are prenumbered and that numerical integrity appears to be maintained.

Also, you may assume that normal review procedures will be followed in such related areas as inventory cut-off, review of aged receivables, uncollectibles, etc.

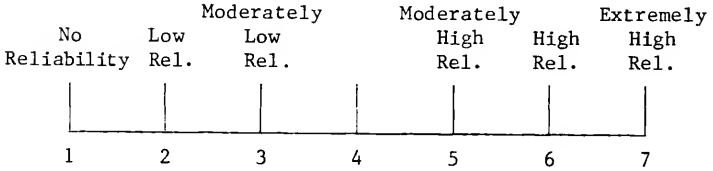
- SCENARIO #1 -

SCENARIO #1 - The following information is common to CASES 1a through 1h. Please disregard any information concerning SCENARIO #2.

The internal control questionnaire shows the following results relative to the existence of the key controls discussed in the information provided previously.

IC-1: YES, procedures exist to control for credit approval.
IC-2: YES, procedures exist to control for shipment authorization.
IC-3: YES, procedures exist to control for the physical shipment of goods.

QUESTION: As a result of all the information you have been provided concerning TMC to this point (including the specific information in this scenario concerning the existence of the three key controls), what degree of reliability would you assign to the given controls with regard to meeting the specific objective of interest? (Circle the appropriate number.)



The following statements require your response as to the most appropriate sample sizes in a preliminary audit program (See NOTE below.) for both compliance and substantive procedures. Using the information provided in the accompanying booklet regarding various sample sizes for compliance and substantive tests, complete the following statements. (You must select a sample size from those provided.)

Compliance Tests - In the selection of sample sizes for the compliance tests, please treat each test individually although the sampling unit (sales invoice packet) will be the same for each test and as a practical matter could be combined.

CT-1, CT-2, & CT-3: Sample size selection and evaluation tables (Tables 1 & 2) are statistically determined.

CT-1: (Test of IC-1): I would recommend that _____ sales invoices be examined to insure that the credit approval procedures have been followed (indicated by the appropriate initials on the supporting customer order.

- This decision was based upon a desired confidence level of _____%. (From Table 1)

CT-2: (Test of IC-2): I would recommend that _____ sales invoices be examined to insure that the shipment authorization procedures have been followed (indicated by the appropriate initials on the sales invoice).

- This decision was based upon a desired confidence level of ____%. (From Table 1)

CT-3: (Test of IC-3): I would recommend that _____ sales invoices be examined to insure that proper control procedures over the physical shipment of goods have been followed (indicated by proper initials on the supporting bill of lading).

- This decision was based upon a desired confidence level of _____%. (From Table 1)

Substantive Test (Sample size Table 3 is statistically derived using a variables sampling approach.)

As a result of the information I have been provided and in conjunction with my planned compliance tests, I would recommend as part of a preliminary audit program the selection of _____ customers from the customer accounts file as of 11/30/78 and that positive confirmations be sent to those selected. (You must select a sample size from Table 3.)

NOTE: In the context of this study, "preliminary audit program" implies an initial audit program that will possibly be adjusted for the substantive test as a result of the evaluation of compliance test results. At that time, a "final audit program" is determined. Also, the substantive test (confirmation of receivables) will be performed on an interim basis as of 11/30/78 with compliance testing performed for the period 1/1/78 - 11/30/78. Appropriate additional procedures will be performed for the month of December.

CASE 1a

- Assume that your recommended sample sizes have been used in the compliance testing of the internal controls that were found to be present.

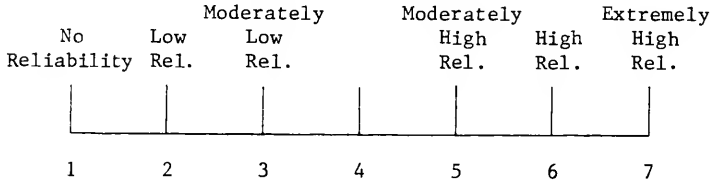
The results of the compliance tests can be summarized as follows:

CT-1: Noncompliance rate = 8%

CT-2: Noncompliance rate = 2%

CT-3: Noncompliance rate = 2%

QUESTION: As a result of all the information you have been provided concerning TMC to this point (including the specific results of your suggested compliance tests provided above), what degree of reliability would you assign to the given controls with regard to meeting the specific objective of interest? (Circle the appropriate number.)



Again, using the information provided in the accompanying booklet regarding various sample sizes for the substantive test, complete the following statement.

Substantive Test (Sample size Table 3 is derived from a statistical variables sampling approach.)

As a result of the information I have been provided and the specific results of the compliance tests, I would recommend as part of a final audit program the selection of _____ customers from the customer accounts file as of 11/30/78 and that positive confirmations be sent to those selected. (You must use a sample size from Table 3.)

CASE 1b

- Assume that your recommended sample sizes have been used in the compliance testing of the internal controls that were found to be present.

The results of the compliance tests can be summarized as follows:

CT-1: Noncompliance rate = 2%

CT-2: Noncompliance rate = 8%

CT-3: Noncompliance rate = 2%

QUESTION: As a result of all the information you have been provided concerning TMC to this point (including the specific results of your suggested compliance tests provided above), what degree of reliability would you assign to the given controls with regard to meeting the specific objective of interest? (Circle the appropriate number.)

No Reliability	Low Rel.	Moderately Low Rel.		Moderately High Rel.	High Rel.	Extremely High Rel.
1	2	3	4	5	6	7

Again, using the information provided in the accompanying booklet regarding various sample sizes for the substantive test, complete the following statement.

Substantive Test (Sample size Table 3 is derived from a statistical variables sampling approach.)

As a result of the information I have been provided and the specific results of the compliance tests, I would recommend as part of a final audit program the selection of _____ customers from the customer accounts file as of 11/30/78 and that positive confirmations be sent to those selected. (You must use a sample size from Table 3.)

CASE 1c

- Assume that your recommended sample sizes have been used in the compliance testing of the internal controls that were found to be present.

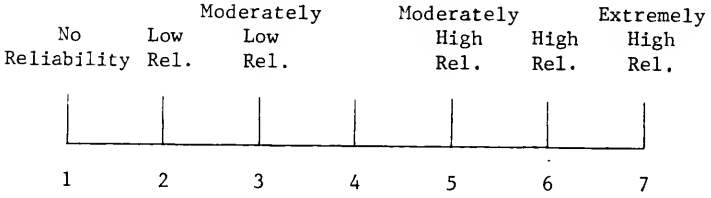
The results of the compliance tests can be summarized as follows:

CT-1: Noncompliance rate = 8%

CT-2: Noncompliance rate = 8%

CT-3: Noncompliance rate = 2%

QUESTION: As a result of all the information you have been provided concerning TMC to this point (including the specific results of your suggested compliance tests provided above), what degree of reliability would you assign to the given controls with regard to meeting the specific objective of interest? (Circle the appropriate number.)



Again, using the information provided in the accompanying booklet regarding various sample sizes for the substantive test, complete the following statement.

Substantive Test (Sample size Table 3 is derived from a statistical variables sampling approach.)

As a result of the information I have been provided and the specific results of the compliance tests, I would recommend as part of a final audit program the selection of _____ customers from the customer accounts file as of 11/30/78 and that positive confirmations be sent to those selected. (You must use a sample size from Table 3.)

CASE 1d

- Assume that your recommended sample sizes have been used in the compliance testing of the internal controls that were found to be present.

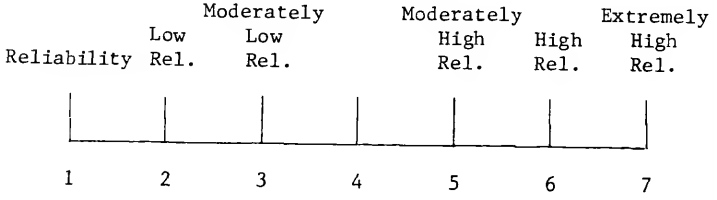
The results of the compliance tests can be summarized as follows:

CT-1: Noncompliance rate = 2%

CT-2: Noncompliance rate = 8%

CT-3: Noncompliance rate = 8%

QUESTION: As a result of all the information you have been provided concerning TMC to this point (including the specific results of your suggested compliance tests provided above), what degree of reliability would you assign to the given controls with regard to meeting the specific objective of interest? (Circle the appropriate number.)



Again, using the information provided in the accompanying booklet regarding various sample sizes for the substantive test, complete the following statement.

Substantive Test (Sample size Table 3 is derived from a statistical variables sampling approach.)

As a result of the information I have been provided and the specific results of the compliance tests, I would recommend as part of a final audit program the selection of _____ customers from the customer accounts file as of 11/30/78 and that positive confirmations be sent to those selected. (You must use a sample size from Table 3.)

CASE 1e

- Assume that your recommended sample sizes have been used in the compliance testing of the internal controls that were found to be present.

The results of the compliance tests can be summarized as follows:

CT-1: Noncompliance rate = 8%

CT-2: Noncompliance rate = 2%

CT-3: Noncompliance rate - 8%

QUESTION: As a result of all the information you have been provided concerning TMC to this point (including the specific results of your suggested compliance tests provided above), what degree of reliability would you assign to the given controls with regard to meeting the specific objective of interest? (Circle the appropriate number.)

No Reliability	Low Rel.	Moderately Low Rel.		Moderately High Rel.	High Rel.	Extremely High Rel.
1	2	3	4	5	6	7

Again, using the information provided in the accompanying booklet regarding various sample sizes for the substantive test, complete the following statement.

Substantive Test (Sample size Table 3 is derived from a statistical variables sampling approach.)

As a result of the information I have been provided and the specific results of the compliance tests, I would recommend as part of a final audit program the selection of _____ customers from the customer accounts file as of 11/30/78 and that positive confirmations be sent to those selected. (You must use a sample size from Table 3.)

CASE 1f

- Assume that your recommended sample sizes have been used in the compliance testing of the internal controls that were found to be present.

The results of the compliance tests can be summarized as follows:

CT-1: Noncompliance rate = 2%

CT-2: Noncompliance rate = 2%

CT-3: Noncompliance rate = 2%

QUESTION: As a result of all the information you have been provided concerning TMC to this point (including the specific results of your suggested compliance tests provided above), what degree of reliability would you assign to the given controls with regard to meeting the specific objective of interest? (Circle the appropriate number.)

No Reliability	Low Rel.	Moderately Low Rel.		Moderately High Rel.	High Rel.	Extremely High Rel.
1	2	3	4	5	6	7

Again, using the information provided in the accompanying booklet regarding various sample sizes for the substantive test, complete the following statement.

Substantive Test (Sample size Table 3 is derived from a statistical variables sampling approach.)

As a result of the information I have been provided and the specific results of the compliance tests, I would recommend as part of a final audit program the selection of _____ customers from the customer accounts file as of 11/30/78 and that positive confirmations be sent to those selected. (You must use a sample size from Table 3.)

CASE 1g

- Assume that your recommended sample sizes have been used in the compliance testing of the internal controls that were found to be present.

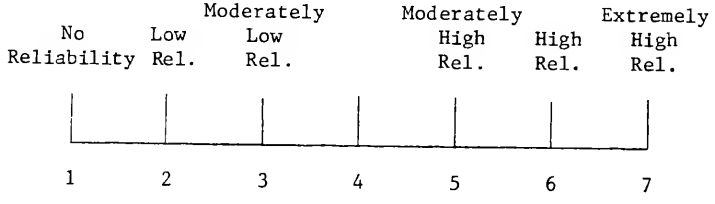
The results of the compliance tests can be summarized as follows:

CT-1: Noncompliance rate = 8%

CT-2: Noncompliance rate = 8%

CT-3: Noncompliance rate = 8%

QUESTION: As a result of all the information you have been provided concerning TMC to this point (including the specific results of your suggested compliance tests provided above), what degree of reliability would you assign to the given controls with regard to meeting the specific objective of interest? (Circle the appropriate number.)



Again, using the information provided in the accompanying booklet regarding various sample sizes for the substantive test, complete the following statement.

Substantive Test (Sample size Table 3 is derived from a statistical variables sampling approach.)

As a result of the information I have been provided and the specific results of the compliance tests, I would recommend as part of a final audit program the selection of _____ customers from the customer accounts file as of 11/30/78 and that positive confirmations be sent to those selected. (You must use a sample size from Table 3.)

CASE 1h

- Assume that your recommended sample sizes have been used in the compliance testing of the internal controls that were found to be present.

The results of the compliance tests can be summarized as follows:

CT-1: Noncompliance rate = 2%

CT-2: Noncompliance rate = 2%

CT-3: Noncompliance rate = 8%

QUESTION: As a result of all the information you have been provided concerning TMC to this point (including the specific results of your suggested compliance tests provided above), what degree of reliability would you assign to the given controls with regard to meeting the specific objective of interest? (Circle the appropriate number.)

No Reliability	Low Rel.	Moderately Low Rel.		Moderately High Rel.	High Rel.	Extremely High Rel.
1	2	3	4	5	6	7

Again, using the information provided in the accompanying booklet regarding various sample sizes for the substantive test, complete the following statement.

Substantive Test (Sample size Table 3 is derived from a statistical variables sampling approach.)

As a result of the information I have been provided and the specific results of the compliance tests, I would recommend as part of a final audit program the selection of _____ customers from the customer accounts file as of 11/30/78 and that positive confirmations be sent to those selected. (You must use a sample size from Table 3.)

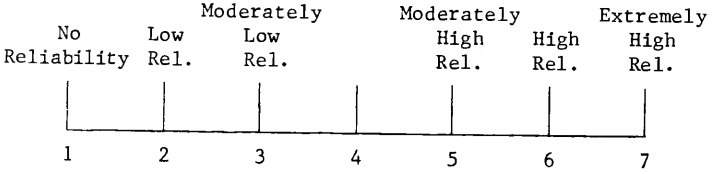
- SCENARIO #2 -

SCENARIO #2 - The following information is common to CASES 2a through 2d. Please disregard any information concerning SCENARIO #1.

The internal control questionnaire shows the following results relative to the existence of the key controls discussed in the information provided previously.

IC-1: YES, procedures exist to control for credit approval.
IC-2: NO, procedures do not exist to control for shipment authorization.
IC-3: YES, procedures exist to control for the physical shipment of goods.

QUESTION: As a result of all the information you have been provided concerning TMC to this point (including the specific information in this scenario concerning the existence of the two key controls), what degree of reliability would you assign to the given controls with regard to meeting the specific objective of interest? (Circle the appropriate number.)



The following statements require your response as to the most appropriate sample sizes in a preliminary audit program (See NOTE below.) for both compliance and substantive procedures. Using the information provided in the accompanying booklet regarding various sample sizes for compliance and substantive tests, complete the following statements. (You must select a sample size from those provided.)

Compliance Tests - In the selection of sample sizes for the compliance tests, please treat each test individually although the sampling unit (sales invoice packet) will be the same for each test and as a practical matter could be combined.

CT-1, CT-2 & CT-3: Sample size selection and evaluation tables (Tables 1 & 2) are statistically determined.

CT-1: (Test of IC-1): I would recommend that _____ sales invoices be examined to insure that the credit approval procedures have been followed (indicated by the appropriate initials on the supporting customer order).

- This decision was based upon a desired confidence level of _____%. (From Table 1)

CT-2: (Test of IC-2): Not appropriate

CT-3: (Test of IC-3): I would recommend that _____ sales invoices be examined to insure that proper control procedures over the physical shipment of goods have been followed (indicated by proper initials on the supporting bill of lading).

- This decision was based upon a desired confidence level of _____%. (From Table 1)

Substantive Test (Sample size Table 3 is statistically derived using a variables sampling approach.)

As a result of the information I have been provided and in conjunction with my planned compliance tests, I would recommend as part of preliminary audit program the selection of _____ customers from the customer accounts file as of 11/30/78 and that positive confirmations be sent to those selected. (You must select a sample size from Table 3.)

NOTE: In the context of this study, "preliminary audit program" implies an initial audit program that will possibly be adjusted for the substantive test as a result of the evaluation of compliance test results. At that time, a "final audit program" is determined. Also, the substantive test (confirmation of receivables) will be performed on an interim basis as of 11/30/78 with compliance testing performed for the period 1/1/78 - 11/30/78. Appropriate additional procedures will be performed for the month of December.

For all decisions, please select the sample sizes you feel are most appropriate prior to consideration of any firm restrictions as to a minimum sample size for a particular compliance or substantive test.

CASE 2a

- Assume that your recommended sample sizes have been used in the compliance testing of the internal controls that were found to be present.

The results of the compliance tests can be summarized as follows:

CT-1: Noncompliance rate = 2%

CT-2: Not appropriate

CT-3: Noncompliance rate = 2%

QUESTION: As a result of all the information you have been provided concerning TMC to this point (including the specific results of your suggested compliance tests provided above), what degree of reliability would you assign to the given controls with regard to meeting the specific objective of interest? (Circle the appropriate number.)

No Reliability	Low Rel.	Moderately Low Rel.		Moderately High Rel.	High Rel.	Extremely High Rel.
1	2	3	4	5	6	7

Again, using the information provided in the accompanying booklet regarding various sample sizes for the substantive test, complete the following statement.

Substantive Test (Sample size Table 3 is derived from a statistical variables sampling approach.)

As a result of the information I have been provided and the specific results of the compliance tests, I would recommend as part of a final audit program the selection of _____ customers from the customer accounts file as of 11/30/78 and that positive confirmations be sent to those selected. (You must use a sample size from Table 3.)

CASE 2b

- Assume that your recommended sample sizes have been used in the compliance testing of the internal controls that were found to be present.

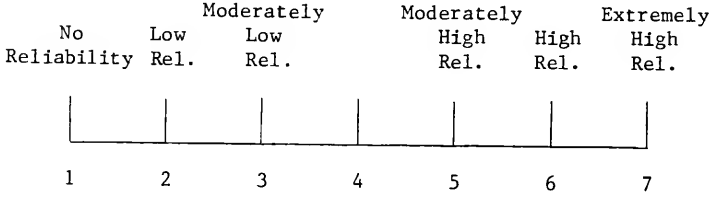
The results of the compliance tests can be summarized as follows:

CT-1: Noncompliance rate = 8%

CT-2: Not appropriate

CT-3: Noncompliance rate = 2%

QUESTION: As a result of all the information you have been provided concerning TMC to this point (including the specific results of your suggested compliance tests provided above), what degree of reliability would you assign to the given controls with regard to meeting the specific objective of interest? (Circle the appropriate number.)



Again, using the information provided in the accompanying booklet regarding various sample sizes for the substantive test, complete the following statement.

Substantive Test (Sample size Table 3 is derived from a statistical variables sampling approach.)

As a result of the information I have been provided and the specific results of the compliance tests, I would recommend as part of a final audit program the selection of _____ customers from the customer accounts file as of 11/30/78 and that positive confirmations be sent to those selected. (You must use a sample size from Table 3.)

CASE 2c

- Assume that your recommended sample sizes have been used in the compliance testing of the internal controls that were found to be present.

The results of the compliance tests can be summarized as follows:

CT-1: Noncompliance rate = 8%

CT-2: Not appropriate

CT-3: Noncompliance rate = 8%

QUESTION: As a result of all the information you have been provided concerning TMC to this point (including the specific results of your suggested compliance tests provided above), what degree of reliability would you assign to the given controls with regard to meeting the specific objective of interest?) (Circle the appropriate number.)

No Reliability	Low Rel.	Moderately Low Rel.		Moderately High Rel.	High Rel.	Extremely High Rel.
1	2	3	4	5	6	7

Again, using the information provided in the accompanying booklet regarding various sample sizes for the substantive test, complete the following statement.

Substantive Test (Sample size Table 3 is derived from a statistical variables sampling approach.)

As a result of the information I have been provided and the specific results of the compliance tests, I would recommend as part of a final audit program the selection of _____ customers from the customer accounts file as of 11/30/78 and that positive confirmations be sent to those selected. (You must use a sample size from Table 3.)

CASE 2d

- Assume that your recommended sample sizes have been used in the compliance testing of the internal controls that were found to be present.

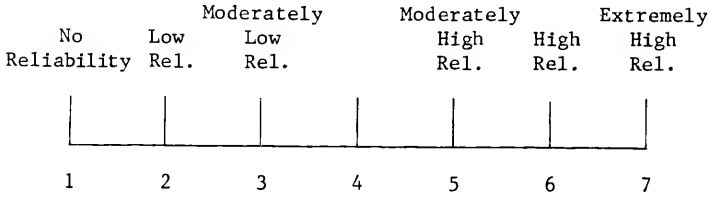
The results of the compliance tests can be summarized as follows:

CT-1: Noncompliance rate = 2%

CT-2: Not appropriate

CT-3: Noncompliance rate = 8%

QUESTION: As a result of all the information you have been provided concerning TMC to this point (including the specific results of your suggested compliance tests provided above), what degree of reliability would you assign to the given controls with regard to meeting the specific objective of interest? (Circle the appropriate number.)



Again, using the information provided in the accompanying booklet regarding various sample sizes for the substantive test, complete the following statement.

Substantive Test (Sample size Table 3 is derived from a statistical variables sampling approach.)

As a result of the information I have been provided and the specific results of the compliance tests, I would recommend as part of a final audit program the selection of _____ customers from the customer accounts file as of 11/30/78 and that positive confirmations be sent to those selected. (You must use a sample size from Table 3.)

- STATISTICAL SAMPLING TABLES -

TABLE 1

- SAMPLE SIZES FOR CT-1 AND CT-2 -

(For One-Sided Upper Precision Limits and Expected Occurrence Rate = 2%)

CONFIDENCE LEVEL = 90%

Sample Size	One-Sided Desired Upper-Precision Limit
25	11.7
50	7.6
75	6.0
100	5.2
125	4.7
150	4.4
175	4.2
200	4.0
225	3.8
250	3.7
275	3.6
300	3.5
350	3.3
400	3.2
475	3.1
550	3.0
650	2.9
800	2.8

CONFIDENCE LEVEL = 95%

Sample Size	One-Sided Desired Upper-Precision Limit
25	14.5
50	9.1
75	7.0
100	6.2
125	5.3
150	5.1
175	4.6
200	4.5
225	4.3
250	4.2
275	4.0
300	3.9
350	3.7
400	3.6
450	3.5
500	3.4
550	3.3
625	3.2
725	3.1
850	3.0
975	2.9

CONFIDENCE LEVEL = 97.5%

Sample Size	One-Sided Desired Upper-Precision Limit
25	15.3
50	10.6
75	8.2
100	7.0
125	6.2
150	5.7
175	5.3
200	5.0
225	4.8
250	4.6
275	4.4
300	4.3
350	4.1
400	3.9
450	3.7
500	3.6
600	3.3
700	3.2
800	3.1
900	3.0
1000	2.9

CONFIDENCE LEVEL = 99%

Sample Size	One-Sided Desired Upper-Precision Limit
25	20.3
50	12.6
75	9.6
100	8.1
125	7.2
150	6.5
175	6.1
200	5.7
225	5.4
250	5.2
275	5.0
300	4.8
350	4.5
400	4.3
450	4.2
500	4.0
600	3.7
700	3.5
800	3.4
900	3.3
1000	3.2

TABLE 2

- EVALUATION OF COMPLIANCE TEST RESULTS (CT-1 & CT-2) -

(Achieved One-Sided Upper Precision Limits)

Confidence Level = 90%

Sample Size	Actual Occurrence Rate (%) 2.0	8.0
25	11.7	19.9
50	7.6	15.4
75	6.0	13.6
100	5.2	12.7
125	4.7	12.1
150	4.4	11.6
175	4.2	11.3
200	4.0	11.0
225	3.8	10.8
250	3.7	10.7
275	3.6	10.5
300	3.5	10.4
350	3.3	10.2
400	3.2	10.0
475	3.1	9.8
550	3.0	9.7
650	2.9	9.5
800	2.8	9.3

Confidence Level = 95%

Sample Size	Actual Occurrence Rate (%) 2.0	8.0
25	14.5	23.1
50	9.1	17.4
75	7.0	15.2
100	6.2	14.0
125	5.3	13.2
150	5.1	12.6
175	4.6	12.2
200	4.5	11.9
225	4.3	11.6
250	4.2	11.4
275	4.0	11.2
300	3.9	11.1
350	3.7	10.8
400	3.6	10.6
450	3.5	10.4
500	3.4	10.3
550	3.3	10.1
625	3.2	10.0
725	3.1	9.8
850	3.0	9.6
975	2.9	9.4

Confidence Level = 97.5%

Sample Size	Actual Occurrence Rate (%) 2.0	8.0
25	15.3	26.2
50	10.6	19.2
75	8.2	16.7
100	7.0	15.1
125	6.2	14.2
150	5.7	13.5
175	5.3	13.0
200	5.0	12.6
225	4.8	12.3
250	4.6	12.1
275	4.4	11.8
300	4.3	11.6
350	4.1	11.3
400	3.9	11.0
450	3.7	10.8
500	3.6	10.6
600	3.3	10.4
700	3.2	10.2
800	3.1	10.0
900	3.0	9.9
1000	2.9	9.7

Confidence Level = 99%

Sample Size	Actual Occurrence Rate (%) 2.0	8.0
25	20.3	29.6
50	12.6	21.5
75	9.6	18.3
100	8.1	16.6
125	7.2	15.5
150	6.5	14.7
175	6.1	14.1
200	5.7	13.6
225	5.4	13.2
250	5.2	12.9
275	5.0	12.6
300	4.8	12.4
350	4.5	12.0
400	4.3	11.7
450	4.2	11.5
500	4.0	11.2
600	3.7	11.0
700	3.5	10.7
800	3.4	10.5
900	3.3	10.3
1000	3.2	10.0

SAMPLE SIZES FOR SUBSTANTIVE TEST

(You must decide on a sample size based upon your choice for the values of Z and M.)*

TABLE 3

M \ CL	(Z=1.65) 90%	(Z=1.96) 95%	(Z=2.24) 97.5%	(Z=2.58) 99%
10,000	390	550	720	959
15,000	175	245	320	426
20,000	100	140	180	240
25,000	65	90	115	153
30,000	43	62	80	107
35,000	30	44	60	78
40,000	25	35	45	61
45,000	20	27	36	47
50,000	15	22	29	38

* The table was developed from the application of a difference estimation approach. Assume any assumptions necessary for the use of the difference estimation approach are met and and table above is derived from the following equation:

$$n = \left(\frac{N \cdot Z \cdot SD}{M} \right)^2$$

where,

n = estimated sample size

N = population size (Fixed at 6,000)

Z = confidence coefficient for various confidence levels (CL)
Z = 1.65 for 90% CL
Z = 1.96 for 95% CL
Z = 2.24 for 97.5% CL
Z = 2.58 for 99% CL

SD = estimated standard deviation of population differences (Fixed at $20, as estimated through appropriate inquiry.)

M = amount judged material for the Sales-Collection Cycle

POSTEXPERIMENT QUESTIONNAIRE

1. Now that you have completed the requirements for the 12 Cases, please allocate 100 points to the three key internal controls in such a way as to indicate the relative importance of each in making your decisions. The more important control should be assigned more points than the less important control, and the total points should equal 100.

   _____ points IC-1 (Credit approval)

   _____ points IC-2 (Shipment authorization)

   _____ points IC-3 (Physical shipment)

   100 TOTAL POINTS

2. Indicate how realistic or representative this task was in comparison to actual audit judgments of a similar nature. (Circle the appropriate number.)

Highly Unrepresentative	Not Representative	Moderately Representative		Very Representative
1	2	3	4	5

3. Indicate what you feel is your degree of familiarity with statistical sampling. (Circle the appropriate number.)

Very Familiar	Familiar	Moderately Familiar	Not Very Familiar	Highly Unfamiliar
1	2	3	4	5

4. What other information would you consider to be important in making similar types of judgment decisions that was not included in this study?

5. Please indicate the extent of your accounting experience.

   Public Accounting ____ years; Industry ____ years; Other ____ years.

   Current Position in Firm _____

   Age _____

6. Please indicate if you are interested in receiving the results of this study.

   ___ NO ___ YES, with names of individuals and/or firms not revealed.

   If YES, please indicate: NAME _____

   ADDRESS _____

Please respond to the following questions by circling the appropriate response. This is not a test; there are no right or wrong answers. It is important that you try to answer each question honestly.

1. Suppose this same experiment was to be repeated on a number of future occasions. How many more times would you be prepared to participate in the experiment under circumstances similar to the present?

No More Times	1 More Time	2 More Times	3 More Times	4 More Times	5 More Times	More Than 5 Times

2. Did your desire to perform well in undertaking a challenging task cause you to try very hard?

Definitely Yes	Yes	Probably Yes	Don't Know	Probably No	No	Definitely No

3. Did you enjoy making the decisions required in the experiment?

Definitely No	No	Probably No	Don't Know	Probably Yes	Yes	Definitely Yes

4. Are you satisfied with your performance in the experiment?

Definitely Yes	Yes	Probably Yes	Don't Know	Probably No	No	Definitely No

5. Did you feel tense during the experiment?

Definitely Yes	Yes	Probably Yes	Don't Know	Probably No	No	Definitely No

6. Did your desire to cooperate with the experimenter cause you to try very hard?

Definitely No	No	Probably No	Don't Know	Probably Yes	Yes	Definitely Yes

7. Did you enjoy the overall experience of participating in the experiment?

Definitely No	No	Probably No	Don't Know	Probably Yes	Yes	Definitely Yes

8. Did your desire to contribute to research knowledge cause you to try very hard?

Definitely Yes	Yes	Probably Yes	Don't Know	Probably No	No	Definitely No

9. Indicate the extent to which you make decisions that are similar to those included in this experiment. (Circle the appropriate number.)

Very Seldom	Seldom	Sometimes	Often	Very Often
1	2	3	4	5

10. Please comment below on any adjustments in your responses to sample size questions had you applied any firm guidelines as to a minimum sample size.

Compliance Tests:

Substantive Tests:

BIOGRAPHICAL SKETCH

Richard Herbert Tabor was born on April 29, 1951, in Oak Ridge, Tennessee. In June 1969 he graduated from Bearden High School, Knoxville, Tennessee.

Mr. Tabor received a Bachelor of Science in Business Administration majoring in accounting in August 1973 from the University of Tennessee. In August 1974 he received a Master of Business Administration majoring in accounting, also from the University of Tennessee. From 1974 until 1976 Mr. Tabor was a Visiting Instructor of Accounting at the University of Evansville, Evansville, Indiana. From June 1976 until the present Mr. Tabor has attended the University of Florida where he has studied for a Doctor of Philosophy degree in business administration (accounting).

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

E. Dan Smith, Chairman
Associate Professor of Accounting

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Douglas A. T. Snowball
Associate Professor of Accounting

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

William F. Messier, Jr.
Assistant Professor of Accounting

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Harry J. Reitz, Jr.
Professor of Management

This dissertation was submitted to the Graduate Faculty of the School of Accounting in the College of Business Administration and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Dean, Graduate School